

A CORRELATION APPROACH
TO TIME-FREQUENCY
REPRESENTATIONS

By

MATTHEW RICHARD PERRY
//

Bachelor of Science in Electrical Engineering
Oklahoma State University
Stillwater, Oklahoma
1984

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1987

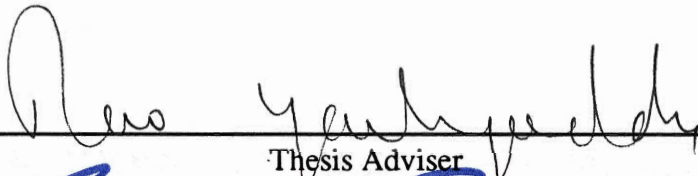
Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 1991

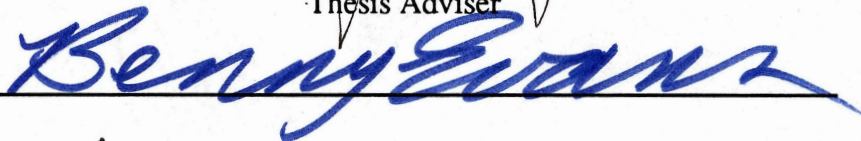
Thesis 1991D

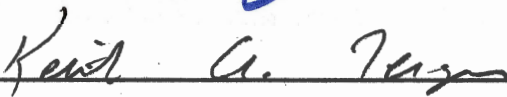
Thesis
1991D
P464c
Cop. 2

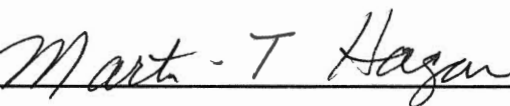
A CORRELATION APPROACH
TO TIME-FREQUENCY
REPRESENTATIONS

Thesis Approved:


Thesis Adviser









Dean of the Graduate College

PREFACE

This thesis investigates the problem of estimating spectral energies from time varying or nonstationary signals. The standard signal processing approach for estimating time varying spectral energies is the spectrogram which assumes signals are short-time stationary. Thus, if a signal is actually highly nonstationary or a signal quickly changes characteristics, the spectrogram produces poor results. As a result, researchers have looked to replace the spectrogram with methods that more effectively estimate time varying spectral energies. Generally, these new techniques use the time-frequency representations with the most popular method being the Wigner distribution. The goal of this thesis is to investigate whether any time-frequency representations exist that are capable of producing better time varying spectral energies when compared to the spectrogram.

Because the Wigner distribution contains anomalies, such as negative values and crossterms, this thesis' goal was expanded to investigate whether features computed on time-frequency representations contain more spectral energy information than corresponding spectrogram features. While investigating the main thesis goal, a new technique called the correlation approach to time-frequency representations was discovered. By changing a time variable transformation and by defining different expected value estimators, the correlation approach was able to compare the periodogram, the power spectral density, the Wigner distribution, the Rihaczek distribution, and Turner's instantaneous power spectrum. Using original correlation approach properties, this thesis showed that time-frequency representations probably are not capable of replacing the spectrogram to estimate time varying spectral energies.

Time-frequency representation features were investigated by implementing a nonstationary indicator whose output at any time stated whether a signal was stationary or nonstationary. Because the nonstationary indicator required a nonlinear mapping between input features and the nonstationary indicator output, a multilayer perceptron neural network implemented the mapping. In addition, the Volterra expansion multilayer perceptron, a new neural network which introduces nonlinear terms into multilayer perceptron's hidden layers, was developed and investigated. Unfortunately, the time-frequency representation's features (instantaneous power, instantaneous frequency, variance, skewness, and kurtosis) did not contain enough time varying spectral energy information to implement the nonstationary indicator. Thus, this thesis has shown that not only will time-frequency representation not replace the spectrogram for estimating time varying spectral energies but the usefulness of time-frequency representation statistical based features is questionable for implementing a nonstationary indicator.

This research could not have been accomplished without the support of the Department of Defense under contract number MDA 904-88-6017 and Sandia National Laboratories under contract number 63-0906. I wish to thank my major adviser, Dr. Rao Yarlagadda, for his guidance and assistance during the course of this research. In addition, I want to extend thanks to my other thesis committee members, Dr. Benny Evans, Dr. Martin Hagan, and Dr. Keith Teague for their time, effort, and support.

Most of all, I would like to thank my wife, Cindy Walker, for her love, constant encouragement, and sound advice.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
Signal Processing	1
Data Classifications	3
Examples of Time Varying Signal Analysis	13
An Electronic Organ	13
Speech Signal Processing	24
Research Goal	29
Research Contributions	30
II. REVIEW OF PAST WORK	34
Introduction	34
Previous Time-Frequency Representations	40
The Spectrogram	40
Page's Instantaneous Power Spectrum	54
Turner's Instantaneous Power Spectrum	57
Levin's Instantaneous Power Spectrum	61
Running Power Spectra and Autocorrelation Functions	64
Rihaczek's Complex Energy Distribution	66
The Wigner Distribution	69
The Wavelet Transformation	70
Cohen's General Time-Frequency Representation	74
Chapter Summary	77
III. THE WIGNER DISTRIBUTION	79
Introduction	79
Wigner Distribution Theory	80
Origins of the Wigner Distribution	80
Wigner Distribution Properties	81
The Discrete Wigner Distribution	100
Aliasing Problems of the Discrete Wigner Distribution	104
Implementing the Discrete Wigner Distribution	105
Wigner Distribution Based Features	106
Basic Distribution Features	107
Signal Processing Applications of the Wigner Distribution	120
Chapter Summary	121
IV. A CORRELATION APPROACH TO TIME-FREQUENCY REPRESENTATIONS	123

Chapter	Page
Introduction	123
Correlation Approach Background	124
Motivation	124
A Brief Random Process Theory Review	126
Correlation Approach to Time-Frequency Representations	138
Chapter Summary	179
 V. THE NONSTATIONARY INDICATOR	 181
Introduction	181
Wigner Distribution Based Features	182
Moment Based Features	183
The Nonstationary Indicator	192
The Multilayer Perceptron	204
The Weighted Linear Combination	204
The Volterra Filter	208
The Nonlinear Combination	210
Arbitrary Mappings	218
The Volterra Expansion Multilayer Perceptron	220
The Multilayer Perceptron and Volterra Expansion Multilayer Perceptron Testing	223
A Nonstationary Indicator Implementation	227
Chapter Summary	242
 VI. SUMMARY AND CONCLUSIONS	 245
Summary	245
Future Work	257
 REFERENCES	 260
 APPENDIX - A REVIEW OF THE HUMAN SPEECH AND HEARING SYSTEMS	 266

LIST OF TABLES

Table	Page
I. A Comparison of Three Time-Frequency Spectral Energies	145
II. A Comparison of Six Time-Frequency Spectral Energies	150
III. A Property Comparison of Six Time-Frequency Spectral Energies	159
IV. A Comparison of Nine Time-Frequency Spectral Energies	175

LIST OF FIGURES

Figure		Page
I.1	Basic Block Diagram of Signal Processing	2
I.2	Signal Classification Examples	9
I.3	Organ Signal Sinusoidal Decomposition	15
I.4	Melody Identification System	16
I.5	Time-Frequency Approach to the Melody Identification System	17
I.6	An Organ Melody Example	19
I.7	The Time-Frequency Domain for the Organ Signal	19
I.8	An Organ Melody Example	20
I.9	The Time-Frequency Domain for the Organ Signal	20
I.10	The Fourier Transformation of an Organ Melody	22
I.11	The Fourier Transformation of an Organ Melody	23
I.12	The Fourier Transformation Based Time-Frequency Representation of the Organ Signal	24
I.13	Speech Signal Sinusoidal Decomposition	26
I.14	A Speech Signal Example	27
I.15	The Fourier Transformation of the Speech Signal	28
II.1	A Time-Invariant Sine Wave	42
II.2	A Rectangular Window and the Rectangular Windowed Sine Wave ...	43
II.3	The Fourier Transform Magnitude of the Sine Wave and Rectangular Windowed Sine Wave	44
II.4	A Time Varying Sine Wave	45
II.5	The Fourier Transform Magnitude of the Time Varying Sine Wave and Rectangular Windowed Time Varying Sine Wave	46

Figure		Page
II.6	Examples of a Rectangular, Hamming, and Gaussian Window	48
II.7	Different Windows Applied to the Sine Wave	49
II.8	Fourier Transformation Magnitude of the Rectangular, Hamming, and Gaussian Windowed Sine Wave	50
II.9	Fourier Transformation Magnitude of the Rectangular, Hamming, and Gaussian Windows	52
III.1	An Example of Time and Frequency Support	85
III.2	The Spectrogram and Wigner Distribution of a Simple Sinusoid	88
III.3	Spectrogram and Wigner Distribution Spectral Energy Estimates of a Simple Sinusoid ($t = 0.5$ secs)	89
III.4	The Spectrogram and Wigner Distribution of a Chirp Signal	91
III.5	Spectrogram and Wigner Distribution Spectral Energy Estimates of a Chirp Signal ($t = 0.5$ secs)	92
III.6	The Spectrogram and Wigner Distribution of Two Sinusoids	93
III.7	Spectrogram and Wigner Distribution Spectral Energy Estimates of Two Sinusoids ($t = 0.5$ and 0.51 secs)	94
III.8	The Spectrogram and Wigner Distribution of Five Signals	97
III.9	Spectrogram and Wigner Distribution Spectral Energy Estimates of Five Signals ($t = 0.5$ and 0.51 secs)	98
III.10	The Organ Signal, the Corresponding Spectrogram, and the Corresponding Wigner Distribution	101
III.11	Spectrogram and Wigner Distribution Spectral Energy Estimates of an Organ Signal ($t = 0.5$ secs)	102
III.12	The Spectrogram Instantaneous Power and the Wigner Distribution Instantaneous Power of a Chirp Signal	110
III.13	The Spectrogram Instantaneous Frequency and the Wigner Distribution Instantaneous Frequency of a Chirp Signal	111
III.14	The Spectrogram Power Spectral Density and the Wigner Distribution Power Spectral Density of a Chirp Signal	112
III.15	The Spectrogram Group Delay and the Wigner Distribution Group Delay for a Chirp Signal	113

Figure		Page
III.16	The Spectrogram Instantaneous Power and the Wigner Distribution Instantaneous Power of an Organ Melody Signal	116
III.17	The Spectrogram Instantaneous Frequency and the Wigner Distribution Instantaneous Frequency of an Organ Melody Signal	117
III.18	The Spectrogram Instantaneous Power and the Wigner Distribution Instantaneous Power of a Speech Signal	118
III.19	The Spectrogram Instantaneous Frequency and the Wigner Distribution Instantaneous Frequency of a Speech Signal	119
IV.1	A Random Process Example	127
V.1	Histograms of the Wigner, Gaussian, and Laplacian Distributions	190
V.2	The Basic Nonstationary Indicator System	193
V.3	The Signal Processing System Which Generates the Nonstationary Indicator	195
V.4	The Ideal Nonstationary Indicator Performance for a Sinusoid	196
V.5	The Ideal Nonstationary Indicator Performance for a Chirp Signal	197
V.6	The Ideal Nonstationary Indicator Performance for Three Concatenated Sinusoids	198
V.7	The Line Used in a Simple Nonstationary Indicator Implementation ...	201
V.8	Nonstationary Indicator Values for the Simple Example	202
V.9	A Linear Combination Network	205
V.10	An Adaptive Linear Combination Network	207
V.11	Nonstationary Indicator Values for the Volterra Expansion Simple Example	210
V.12	An Adaptive Nonlinear Combination Network	212
V.13	Graphs of the Sigmoid Nonlinear Function and the Hyperbolic Tangent Nonlinear Function	213
V.14	A Multilayer Perceptron	215
V.15	A Multilayer Perceptron with One Hidden Layer and One Output	220
V.16	The Nonlinear Function Average Training Error for a 3-16-1 Volterra Expansion Multilayer Perceptron, a 3-16-1 Multilayer Perceptron, and a 3-44-1 Multilayer Perceptron	226

Figure		Page
V.17	A Test Signal Consisting of Sinusoids and Chirp Signals and the Ideal Nonstationary Indicator	233
V.18	A Test Signal Consisting of Sinusoids and Chirp Signals and the Ideal Nonstationary Indicator	234
V.19	Average Training Errors for a 5-20-1 Multilayer Perceptron and a 5-16-1 Volterra Expansion Multilayer Perceptron	235
V.20	The Nonstationary Indicator Output of a 5-20-1 Multilayer Perceptron and a 5-16-1 Volterra Expansion Multilayer Perceptron on the Trained Data	237
V.21	The Nonstationary Indicator Output of a 5-20-1 Multilayer Perceptron and a 5-16-1 Volterra Expansion Multilayer Perceptron on the Trained Data	238
V.22	The Nonstationary Indicator Output of a 5-20-1 Multilayer Perceptron and a 5-16-1 Volterra Expansion Multilayer Perceptron on Untrained Data	239
V.23	The Nonstationary Indicator Output of a 5-20-1 Multilayer Perceptron and a 5-16-1 Volterra Expansion Multilayer Perceptron on Untrained Data	240

CHAPTER I

INTRODUCTION

Signal Processing

Signal processing is the science of studying, understanding, and generating systems and signals. The block diagram in Figure I.1 helps illustrate this signal processing definition by representing the system with a black box and representing signals as black box input and output. In the block diagram, the input, the output, or the system may be either known or unknown. In most cases however, signal processing techniques typically analyze the unknown quantities of Figure I.1. As such, signal processing describes a wide range of concepts.

In more technical terms, signal processing and the block box model can be described mathematically as a mapping

$$f : A \rightarrow B, \quad (\text{I.1})$$

where f represents the signal processing system or black box, A represents an abstract set of inputs, and B represents an abstract set of outputs. Signal processing then defines methods by which f , A , and B are estimated, generated, or filtered based on known information concerning f , A , and B . As an example of a common application in signal processing, signal filtering removes unwanted noise from signals. Given a noisy signal A , signal processing defines a system f so that B contains less noise than the signal A . A second signal processing example is system identification where signals A and B estimate the system f . A third, and more abstract, signal processing example is pattern recognition. Pattern recognition identifies abstract patterns from a given input. Thus, A is a

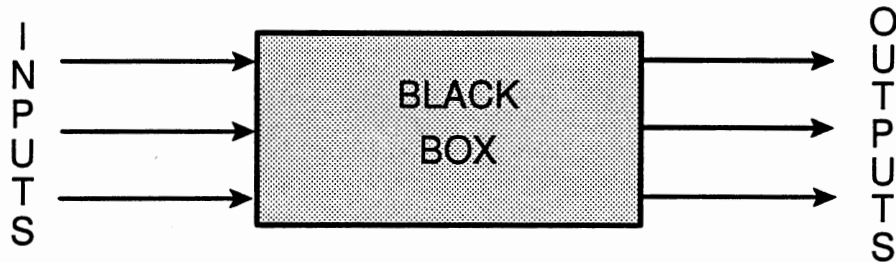


Figure I.1 Basic Block Diagram of Signal Processing

signal, B is a set of abstract patterns, and f transforms the signal to the abstract pattern.

The system f is complex because the signal typically requires filtering and feature extraction. Additionally, the function must map the input signals to some abstract set of patterns.

These three signal processing examples show the potential of signal processing. Numerous other examples that modify the known and the unknown signals and systems exist and span numerous fields. Thus, signal processing is a broad topic with many useful applications. Signal processing has been an active research field for hundreds of years and especially active during the past twenty years. The recent increase in research activity resulted from the wide availability of computers that allow the simulating and the testing of signal processing algorithms. However, much signal processing work remains since real world problems require simplifying assumptions in applying problems to the abstract signal processing model. As an example, most signal processing techniques assume that systems do not vary in time. However, many real world systems' characteristics change over time. Signals developed from systems that change over time are called

time varying signals. One example of a time varying signal is speech. Because each unique speech sound requires a different anatomical configuration, a system that changes over time produces speech signals [Rab78]. Therefore, applying signal processing techniques to speech signals requires algorithms that are capable of processing time varying signals. Another common simplifying assumption is that linear models characterize a system. However, this assumption typically fails since most real world systems contain some nonlinear components.

Therefore, advancing signal processing requires investigating more accurate descriptions of the signal processing black box model. This thesis attempts to enhance signal processing techniques for time varying signals. Because the Fourier transformation is so widely used in signal processing, this thesis specifically investigates methods of applying Fourier transformation ideas to time varying signals. Before delving into the investigation of time varying signals, a review of basic signal processing terminology is presented.

Data Classifications

The word signal processing has been defined and described abstractly without stating what the word signal actually means. For this thesis, a *signal* is a single valued function of one variable. The function is not necessarily continuous, differentiable, or mathematically "nice" in any way. However, most signals dealt with in signal processing do possess many convenient properties for analysis. Probably the most common signal is the real valued function defined by

$$\text{signal} = f(t), \quad t \in \mathfrak{R}, \quad f(t) \in \mathfrak{R}, \quad (\text{I.2})$$

where t represents time, space, or some other variable and $f(t)$ represents the function's value or the signal value. Here, t is used to represent time for simplicity. A simple example of (I.2) is $f(t) = \sin(\omega t)$ where ω is a radian frequency.

Signals can be classified in many different ways. First, if the preimage, t in (I.2), is real valued then a signal is said to be *continuously sampled*. Conversely, a signal is *discretely sampled* if its preimage is a subset of the integers. Thus, a *discretely sampled signal* contains at most a countable infinite number of values; whereas, *continuous sampled signals* contain an uncountable infinite collection of values. The main difference between these two classifications is that discretely sampled signals correspond to signals capable of residing in computer memory, whereas continuously sampled signals correspond to naturally occurring signals. The reader should note that a continuously sampled signal is typically called a continuous signal in signal processing literature, even though the signal may not be continuous in the mathematical sense. Likewise, discretely sampled signals are often referred to as discrete signals, although the signal may be a subset of the real numbers.

Another major signal classification indicates how function values behave. If the function values, $f(t)$ in (I.2), are continuous in the mathematical sense, then the signal is a *continuous valued* signal. Conversely, if function values, $f(t)$ in (I.2), are integer subsets then the signal is termed *digital valued*. The main difference between these two classifications in electrical engineering corresponds to the hardware implementing method. Analog electronics, consisting of continuous valued functions, implement continuous valued signals. Alternately, digital electronics, corresponding to computers, typically implement digital valued signals. Thus, because computers can store digital signals, simulating signal processing procedures over digital signals is easily accomplished using computers. Similarly, because mathematical concepts such as integration, differentiation, etc. are based on continuous functions, theoretically evaluating signal processing algorithms is easier using continuous signals.

For the signal classifications discussed so far (continuous valued continuous signals, continuous valued discrete signals, digital valued continuous signals, and digital

valued discrete signals), only two classifications predominate in signal processing: continuous valued continuous signals (called analog signals) and digital valued discrete signals (called digital signals). Analog and digital signals dominate signal processing for two reasons. First, analog signals, as stated previously, are found readily in nature and are subject to numerous mathematical operations while digital signals implement easily on computers. Second, a correspondence between these two signal types exists using Nyquist's Sampling Theorem [Sch80]. This theorem states that if an analog signal is sampled at greater than twice its bandwidth, then the digital signal completely describes the continuous signal. Using this digital signal, the analog signal can be completely recovered using an interpolation formula. Thus, theoretical evaluations performed upon analog signals apply directly to digital signal representations of the analog signals. Similarly, computer based digital signal simulations correspond to simulating analog signals. Hence, in this thesis, theoretical work uses analog signals while simulations and examples use digital signals.

Now that different versions of the basic signal definition, (I.2), have been discussed, other signal characteristics or signal classifications based on how the signal behaves is reviewed. Probably the most significant signal characteristic relies on whether a signal is predictable or unpredictable. Predictable signals are called *deterministic signals*. A deterministic signal can be thought of as a function for which an expression can be written. Examples include sinusoidal and constant functions. Digital signals that have already been acquired in a signal processing system are included in deterministic signals. The reason for including digital signals is that since the values are known, the signals are predictable. The opposite of deterministic signals is *random signals*. A random signal is all or partially unpredictable. Thus, an exact expression is impossible to write down. All real world signals are random due to the presence of noise.

In terms of signal processing analysis, deterministic and random signals are handled differently. Since some equation or actual signal data is available for deterministic data, exact theoretical analysis is possible. Random signals do not fit well with direct theoretical analysis since the exact signal values are never known. Because exact signal values are never known, a probabilistic structure describing the random data is used. Even with probability information, theoretical analysis of random data is not easy. Thus, electrical engineers typically perform an initial analysis assuming a signal is deterministic and then analyze how a signal processing procedure performs for the ideal case. Once the ideal case is investigated, effects of randomness are studied using techniques such as sensitivity analysis and others.

As previously mentioned, most signal processing methods assume a system is linear and does not vary with time. Why are these two things assumed and what do these terms mean? First, a system is *linear* if and only if

$$y = f(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 f(x_1) + \alpha_2 f(x_2) = \alpha_1 y_1 + \alpha_2 y_2$$

where

$$y_i = f(x_i), \quad i = 1, 2 \quad (\text{I.3})$$

for any inputs x_1 and x_2 and for any $\alpha_1, \alpha_2 \in \mathfrak{R}$. Otherwise, the system is called *nonlinear*. Assuming a linear system allows writing the system response as

$$y(t) = \int_{-\infty}^{\infty} h(t, \tau) x(\tau) d\tau, \quad (\text{I.4})$$

where $h(t, \tau)$ is called the *impulse response* of the system. If $h(t, \tau)$ is known for all τ , then the output, $y(t)$, can be computed for any input, $x(t)$ [Che84]. Computing the output based on the input allows much *a priori* system analysis.

The only problem with (I.4) is that $h(t, \tau)$ is a function of time t and τ . A nicer situation occurs if the impulse response depends on only one variable. Obtaining an

impulse response of one variable is where the time-invariant assumption enters. A system is *time-invariant* if system characteristics do not change with time. Relating this definition with (I.4), a system is time-invariant if

$$h(t, \tau) = h(t - \tau), \quad (\text{I.5})$$

so that

$$y(t) = \int_{-\infty}^{\infty} h(t - \tau)x(\tau)d\tau = h(t) \otimes x(t). \quad (\text{I.6})$$

Equation (I.6) is called the *convolution integral*. The convolution integral is critical in signal processing because (I.6) can be described using the Fourier transformation

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft} dt \quad (\text{I.7})$$

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{i2\pi ft} df \quad (\text{I.8})$$

and using

$$Y(f) = H(f) \cdot X(f), \quad (\text{I.9})$$

where $Y(f)$, $H(f)$, and $X(f)$ are the Fourier transformations of $y(t)$, $h(t)$, and $x(t)$, respectively. Therefore, a linear and time-invariant system can be described quite easily because (I.9) is a very simple operation compared to the convolution integral, (I.6).

A system that changes characteristics over time is called *time-variant* or *time varying*. Either time varying systems or nonlinear systems disallow using (I.9) for signal processing analysis. As an example of a nonlinear system, if $y(t) = x^2(t)$ then its Fourier transformation is $Y(f) = X(f) \otimes X(f)$ which is much more complicated than (I.9). Thus, not using (I.9) makes developing signal processing procedures difficult. Therefore, methods are sought to analyze and to understand systems that are either nonlinear or time varying. Because the topic of nonlinear systems is more widely researched than time

varying systems, this thesis investigates methods of analyzing time varying systems. Ideally, techniques are sought which allow using a form of (I.9) which is valid for time varying signals. Thus, this thesis searches for a time varying version of the Fourier transformation.

The reader should note that time-invariant systems are sometimes called stationary systems while time varying systems are typically called nonstationary systems. Although these terms have similar meanings, stationary and nonstationary are usually reserved for random signals. Thus, for this thesis, time varying systems include both deterministic and random signals while nonstationary signals are defined later in this section.

As is usually the case, examples can clarify definitions. Thus, Figure I.2 attempts to distinguish the difference between a time-invariant signal and a time varying signal. The first graph in Figure I.2 represents a sinusoid

$$x(t) = \sin(2\pi ft), \quad (\text{I.10})$$

where f equals 5 Hz. Since the frequency content of (I.10) does not change over time, the signal is time-invariant. The next graph shows a function called a chirp signal described by

$$x(t) = \sin(2\pi ft^2), \quad (\text{I.11})$$

where f equals 20 Hz. This signal appears to be the same as the time-invariant sinusoid except that the time variable is squared. Rewriting (I.11) as

$$x(t) = \sin(2\pi(ft)t) \quad (\text{I.12})$$

indicates that the frequency of the sinusoid frequency increases linearly from 0.0 Hz to 20.0 Hz for times between 0.0 seconds and 1.0 seconds. Thus, the Fourier spectrum describing (I.12) varies with time. Hence, the chirp signal is a time varying signal. Notice that even though the chirp signal is simple to write down and appears normal,

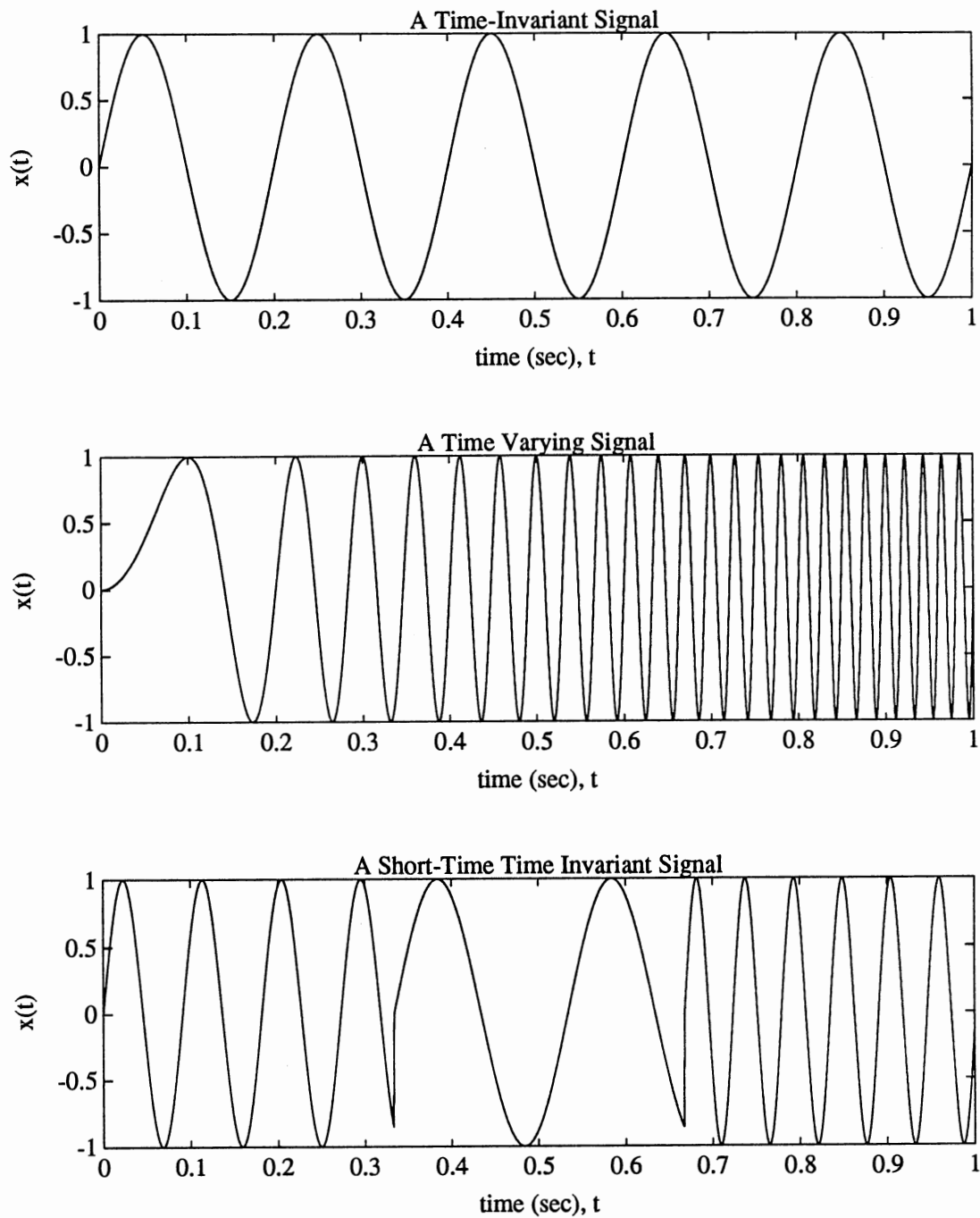


Figure I.2 Signal Classification Examples

the chirp signal is classified as time varying. The third example covers a middle ground between a time-invariant signal and a time varying signal. The third graph in Figure I.2 illustrates a concatenation of three different sinusoids. This signal is given by

$$x(t) = \begin{cases} \sin(2\pi f_1 t) & \forall t \in [0.0, .333] \\ \sin(2\pi f_2 t) & \forall t \in [.333, .667] \\ \sin(2\pi f_3 t) & \forall t \in [.667, 1.0] \end{cases}, \quad (\text{I.13})$$

where $f_1 = 11$ Hz, $f_2 = 5$ Hz, and $f_3 = 18$ Hz. In the strict sense of the definitions, this signal is time varying because the spectrum describing the signal changes twice. However, this signal is remarkably different than the chirp signal which changes spectrum an infinite number of times. Additionally, the concatenated signal is time-invariant in intervals. Because this signal type represents a broad number of real world signals, a new classification, *short-time time-invariant*, signifies these signals that are mainly time-invariant. This definition allows using time-invariant signal processing algorithms in cases of short-time time-invariant signals. However, these techniques, which depend on a time-invariant signal, are destined to fail during the spectral changes. Failure during spectral changes or during *transition regions* justifies investigating signal processing techniques specifically designed for time varying signals. Transitions regions are very important in many signal processing applications. For example, in speech signal processing, transitions occur from one phoneme to another [Rab78].

The definitions of time varying signals and time-invariant signals cover both deterministic and random data. Random signal literature has produced other terms based solely on probabilistic quantities. The random signal equivalent to a time-invariant signal, called a *stationary signal*, contains statistical quantities which do not vary with time. Notice that in this strict definition, time differences are not allowed. Including time differences into the definition produces random signals called *wide-sense stationary*. As already indicated, wide-sense stationary signals depend on time differences but not on

time explicitly. The random signal equivalent to time varying signals are called *nonstationary* signals. Nonstationary signals contain some statistical quantity that depends on time.

Because most real world signals are random, the concepts of stationary and nonstationary signals are further developed. First, random signals cannot be written down directly in the form $x(t)$. thus, a random signal is described by a *random process* $\{X_k(t), k \in Z, t \in \mathfrak{R}\}$ where Z represents the positive integers; for a particular t , $X_k(t)$ is a *random variable*; for a particular k , $X_k(t)$ is a *sample function*; and for a specific t and k , $X_k(t)$ is a signal value. Unfortunately, a signal value is only one of a countable infinite possible values that occurs for a particular time. The random variable probability density function characterizes the distribution of these values. Thus, knowledge of the probability density function at a particular time is somewhat analogous to knowing the actual signal values of deterministic signals. Therefore, when dealing with random signals, the most useful information is a distribution of possible signal values. However, the situation can worsen when the random variable probability density function is not known; in such case, estimating probability density moments, such as the mean and the variance, is the best tact. These probability density moments are estimated using *ensemble averages* (averages taken over all random variables) at a particular time. Thus, random signals are very difficult to work with because only the ensemble average values for a random signal are known instead of actual signal values. A problem exists, however, when dealing with random signals in real world signal processing systems. Typically, only one sample function is available for processing. Thus, only one realization of the random process is available for computing ensemble averages. Because an ensemble average over one sample is impractical, engineers typically replace ensemble averages with time averages. A *time average* is an average taken over the time variable of a sample function. Should the time averages equal the ensemble averages, the random process is called *ergodic*. Because time averages do not depend on time, ergodic processes are automatically sta-

tionary. However, stationary random processes may not be ergodic. Knowing when stationary signals are ergodic is difficult to determine except when the random process is Gaussian distributed since finite moments guarantee an ergodic signal [Gar86]. Thus, even though engineers compute time averages in place of ensemble averages, the practice does not always produce correct results.

The discussion of time averages and ensemble averages indicates the problems faced by engineers who deal with nonstationary signals. One problem is that time averages do not necessarily yield ensemble averages. Thus, even though time averages compute easily, time averages may not contain any information regarding the underlying random process. Currently, considerable research exists into the analysis of nonstationary signals, and this thesis adds to the topic.

Although time varying signals occur naturally, not many time varying signal techniques have been incorporated into new signal processing algorithms. If new signal processing algorithms are developed for time varying signals, an obvious signal processing technique would alter the Fourier transformation since many signal processing algorithms rely on Fourier information. Fourier analysis transforms or decomposes a signal into complex exponentials (the only eigenfunctions of linear systems). Because complex exponentials are the only eigenfunctions for linear systems, the output of a linear system with a complex exponential input is itself a complex exponential. Thus, observing the input and output signals completely characterize a system.

The second reason that signal processing relies so heavily on spectral analysis is that many naturally occurring signals are periodic (at least on a short time basis). Thus, sinusoidal analysis becomes an efficient method of decomposing a signal. Additionally, the source of many real world signals can most easily be described in terms of harmonic oscillation. The next section of this thesis presents two unsolved signal processing problems. Although these two unsolved problems will seem quite simple, the problems are

difficult because the problems contain short-time time-invariant signals. These seemingly simple examples will indicate how additional research into time varying and nonstationary signal processing can yield fruitful returns.

Examples of Time Varying Signal Analysis

This section describes two examples which relate current spectral analysis techniques to problems associated with time varying signals. The first example introduces the problem of identifying a simple melody from a constrained organ. The organ example exemplifies the fact that analyzing even simple signals is difficult. The second example extends the simple organ example to the more real world problem of speech signal processing where potential applications identify speech signals, enhance speech signals, and code speech signals.

An Electronic Organ

The first example identifies simple melodies generated by a constrained electronic organ. The example organ is similar to cheap toy organs found in many toy stores. As such, the example organ output contains only one note at a time. Although this organ does not generate arbitrary sounds, a single note system reduces the signal complexity to a sinusoid of a particular frequency. Thus, the organ produces one distinct sinusoid for any one time. Additionally, the organ's output sinusoid is fixed so its peak amplitude equals one. Forcing a constant peak value insures that all organ sounds contain equal power or loudness. Therefore, the organ can, at any time t , produce the organ signal

$$x(t) = \sin(2\pi f_t t + \phi_t), \quad (\text{I.14})$$

where f_t corresponds to the appropriate note at time t while ϕ_t is the phase at time t . Note that the organ signal, (I.14), is very simplistic; thus, the organ output should allow simple identifying signal processing procedures. The example melodies are generated by a

musician or a performer using a musical score. The musical score contains melody notes and note playing times. Thus, the performer plays a melody based on a time versus note representation, or a time versus (frequency and phase) representation.

An organ produced signal can be pictured as a concatenation of sinusoids because each note is a constant powered sinusoid. Figure I.3 illustrates a graph of an arbitrarily produced organ signal by showing the digital signal equivalent of a continuous organ signal for three time samples. Each time sample contains two parameters: frequency and phase. Many melodies have locations where notes remain constant for many time samples, so the three sinusoids in Figure I.3 could possess equal parameters. However, the possibility exists that all three sinusoids contain different parameters.

Signal processing begins analyzing an organ melody by acquiring the organ signal for processing. Signal acquisition requires either recording the signal on analog tape with the aid of a speaker output and a microphone, acquiring the signal digitally from a speaker output and microphone, or directly connecting the organ to a signal acquiring system. Regardless of the acquiring method, the signal output is random due to induced noise. Thus, signals acquired in real world systems are random, even if the original signals are deterministic. The process of estimating the melody from the acquired organ signal is called pattern recognition and is an integral part of signal processing. The entire system of performing and acquiring the melody signal is illustrated by Figure I.4.

The question remains as to how signal processing reliably estimates an organ melody. Investigating a signal's origin typically produces a good starting point for signal processing procedures. The starting point for this example would be the musical score that represents the melody in the time versus (note) domain. Thus, investigating a signal's sinusoidal content at a particular time should allow for robust melody estimating procedures. Figure I.5 illustrates this idea by modifying Figure I.4 in the appropriate places.

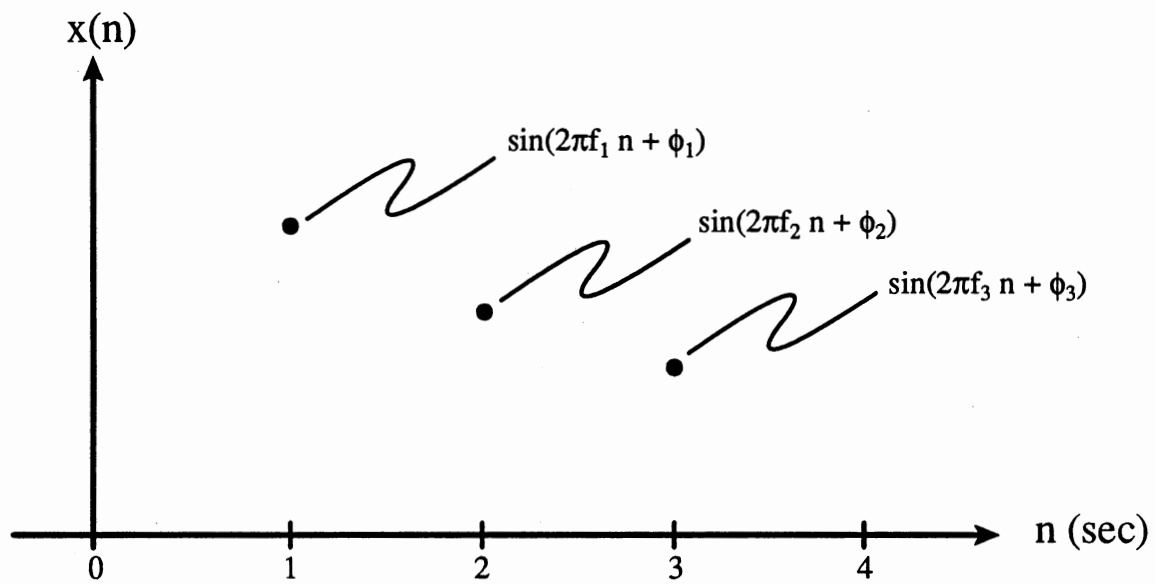


Figure I.3 Organ Signal Sinusoidal Decomposition

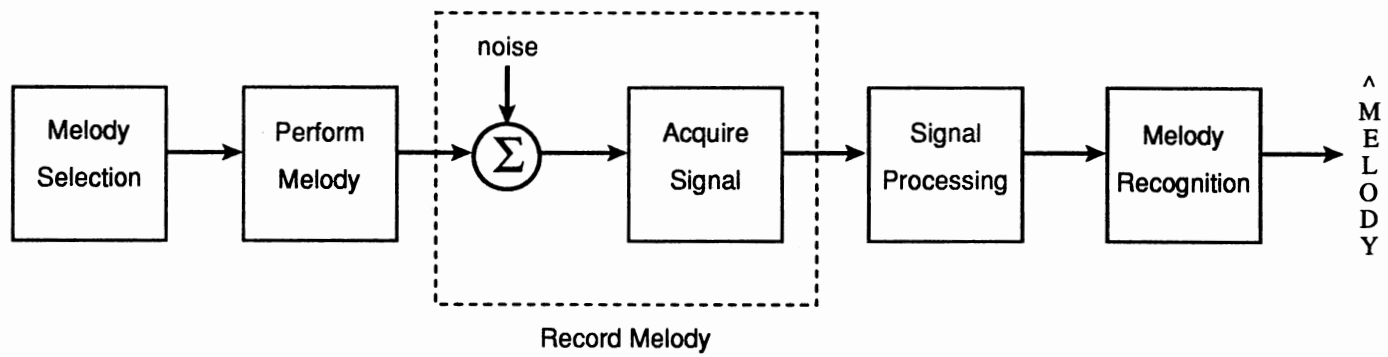


Figure I.4 Melody Identification System

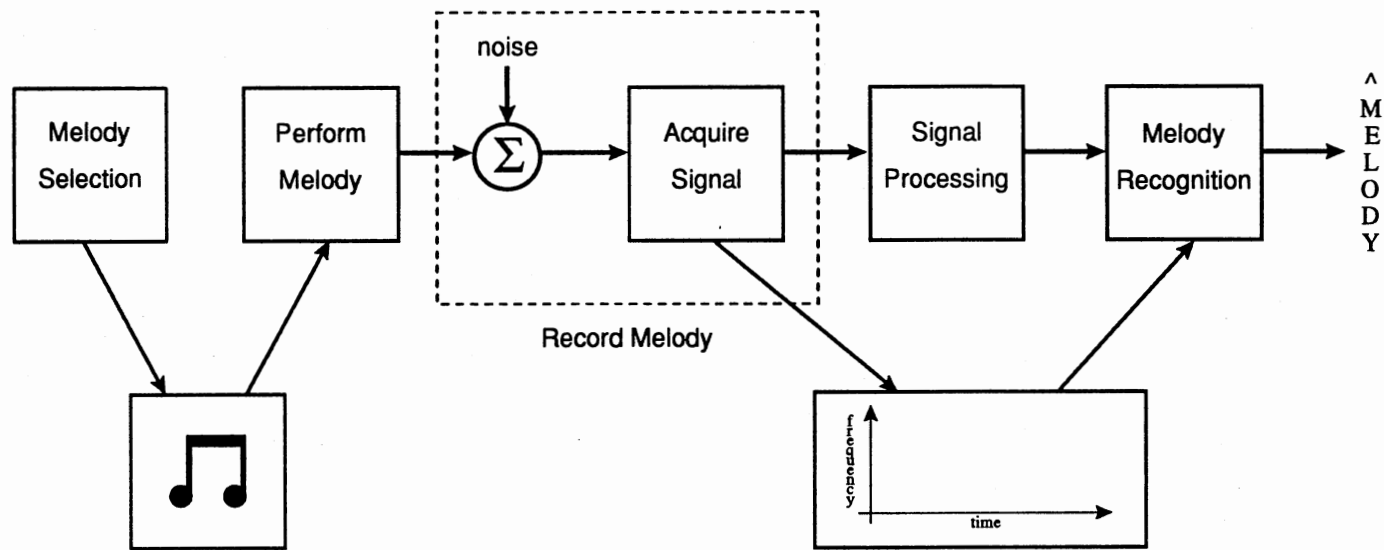


Figure I.5 Time-Frequency Approach to the Melody Identification System

Additionally, the organ constrains the output to a constant power single sinusoid. Thus, the organ signal characteristics and the original melody source, namely the musical score, indicate that estimating the sinusoid at each time may produce robust melody estimates. Estimating the sinusoid at each time reduces to estimating the parameters of (I.14) (f_i and ϕ_i) for each time of interest. However, since the human ear is rather phase insensitive (see Appendix), the phase term, ϕ_i , is irrelevant for estimating the melody. Thus, only the sinusoid frequency for a particular time is estimated. Because the phase term is irrelevant, the time versus (frequency and phase) melody representation reduces to a time-frequency representation. Therefore, estimating a melody is possible by estimating the melody's time-frequency representation.

To get a feeling for time-frequency representations, two different but related melodies are analyzed. The first melody, shown in Figure I.6, is defined by

$$x(t) = \begin{cases} \sin(2\pi f_1 t + \phi_1) & \forall t \in [0.0, .333] \\ \sin(2\pi f_2 t + \phi_2) & \forall t \in [.333, .667] \\ \sin(2\pi f_3 t + \phi_3) & \forall t \in [.667, 1.0] \end{cases}, \quad (\text{I.15})$$

where $f_1 = 24.3$ Hz, $f_2 = 6.1$ Hz, $f_3 = 87.5$ Hz, $\phi_1 = 7\pi/8$ radians, $\phi_2 = \pi/4$ radians, and $\phi_3 = \pi/2$ radians. This organ signal concatenates three different sinusoids. Ideally, the contour graph in Figure I.7 represents the signal's time-frequency content. On the contour graph, the black lines represent the third dimension of the domain, or the spectral energy. Thus, the presence of a sinusoid at a particular time is represented by a black dot in the appropriate time and frequency location. Therefore, Figure I.7 indicates the existence of three sinusoids occurring at different times.

The second organ melody is obtained by shifting the first melody. This organ signal is shown in Figure I.8 and is described by (I.15) except $f_1 = 6.1$ Hz, $f_2 = 87.5$ Hz, $f_3 = 24.3$ Hz, $\phi_1 = \pi/4$ radians, $\phi_2 = \pi/2$ radians, and $\phi_3 = 7\pi/8$ radians. Like the first organ signal, this organ signal concatenates three sinusoids. The ideal time-frequency

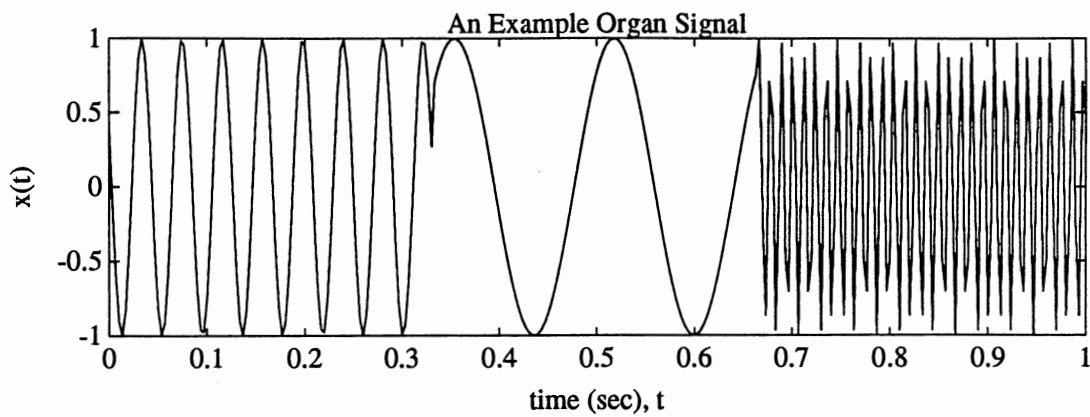


Figure I.6 An Organ Melody Example

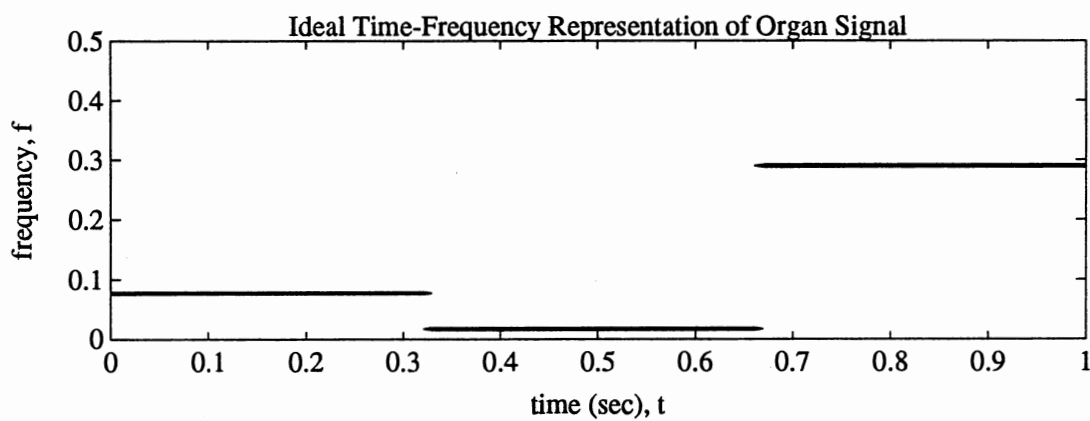


Figure I.7 The Time-Frequency Domain for the Organ Signal

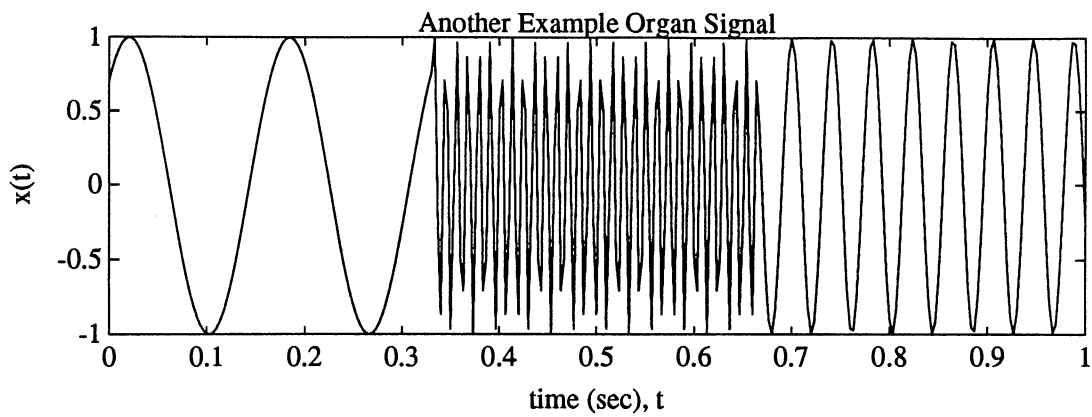


Figure I.8 An Organ Melody Example

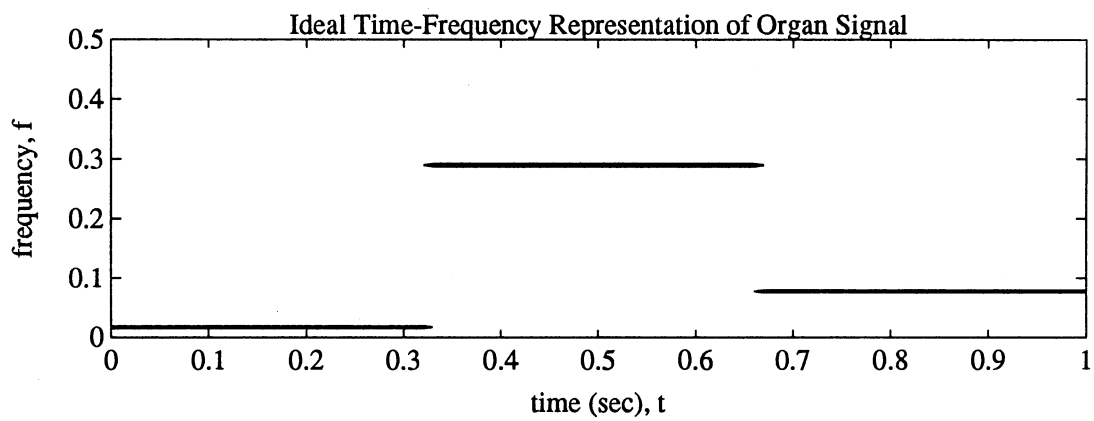


Figure I.9 The Time-Frequency Domain for the Organ Signal

representation for this organ signal is given by the contour graph in Figure I.9. Again, the ideal time-frequency domain indicates the presence of three sinusoids occurring at different times. Notice the similarity in the time-frequency representations of these two example signals. This similarity indicates that the two signals are just time shuffled versions of each other.

The question remains as to how signal processing generates these ideal time-frequency representations. The simplest method uses the standard technique for estimating spectrums, the Fourier transformation defined in (I.7). The Fourier transformations of the two example organ signals are shown in Figures I.10 and I.11, respectively. Note the similarity of the magnitude spectra for the two cases and the difference in their phase spectra. Because the Fourier transformation produces complex results, the spectral estimates are displayed using the magnitude and phase signals.

Although the organ signal's Fourier transformation magnitude is not a function of time, a time-frequency representation can be generated by using the same spectral estimates for each time value. This procedure yields a Fourier transformation time-frequency representation. Because the magnitudes of the two organ signals are the same, the contour graph in Figure I.12 represents the time-frequency representation of both signals. Note that this time-frequency representation does not provide any temporal information. The lack of temporal information is evident by comparing Figure I.12 with the ideal graphs in Figures I.7 and I.9.

Referring back to Figures I.10 and I.11, the temporal information lost in the contour graph in Figure I.12 exists in the phase signals. Since phase signals are not directly present in the time-frequency representation given in Figure I.12, how can the temporal information be included into time-frequency representations? One method would place the phase information in an entirely different signal. However, if phase information is desired directly in the time-frequency representations, then extending the Fourier transformation in this manner is not sufficient.

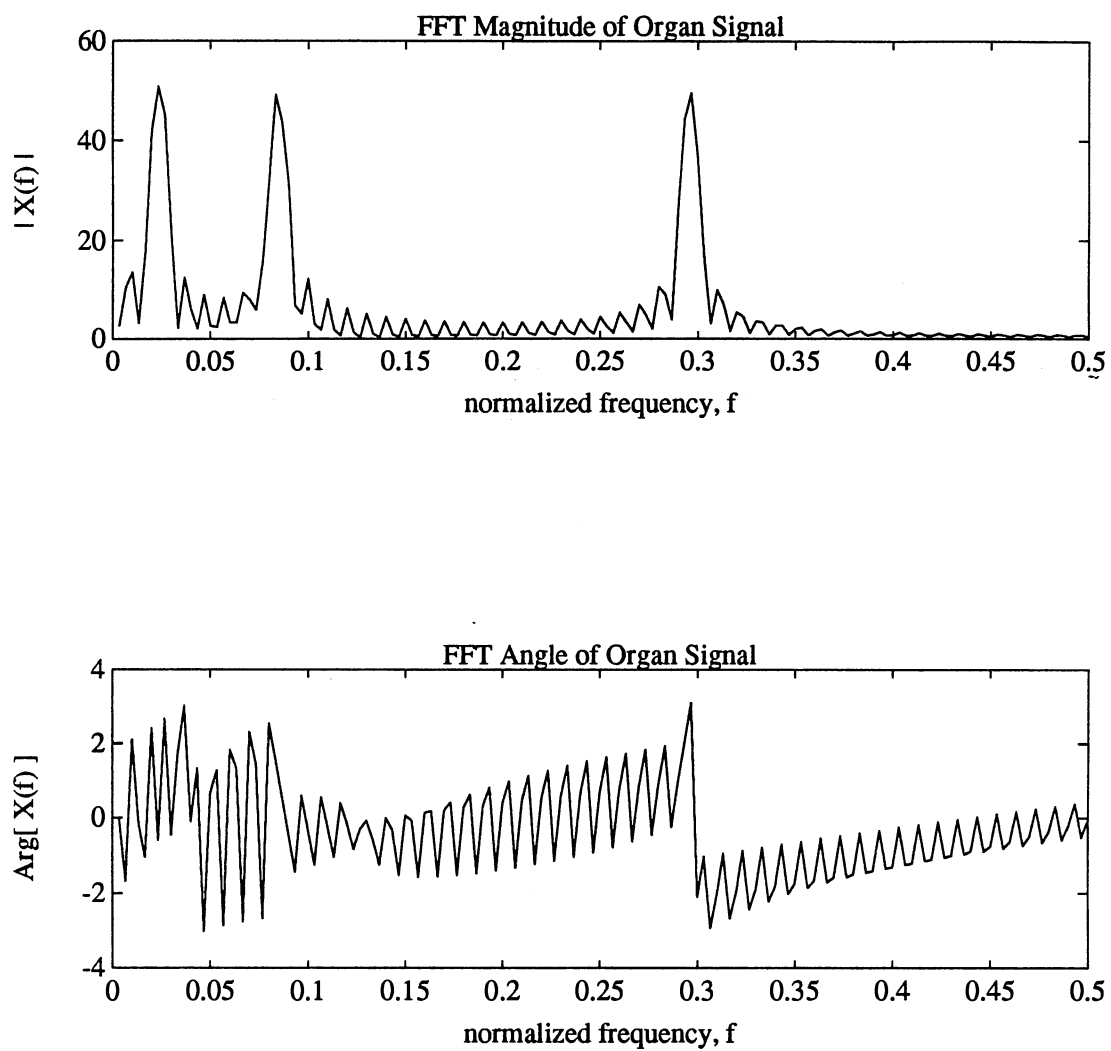


Figure I.10 The Fourier Transformation of an Organ Melody

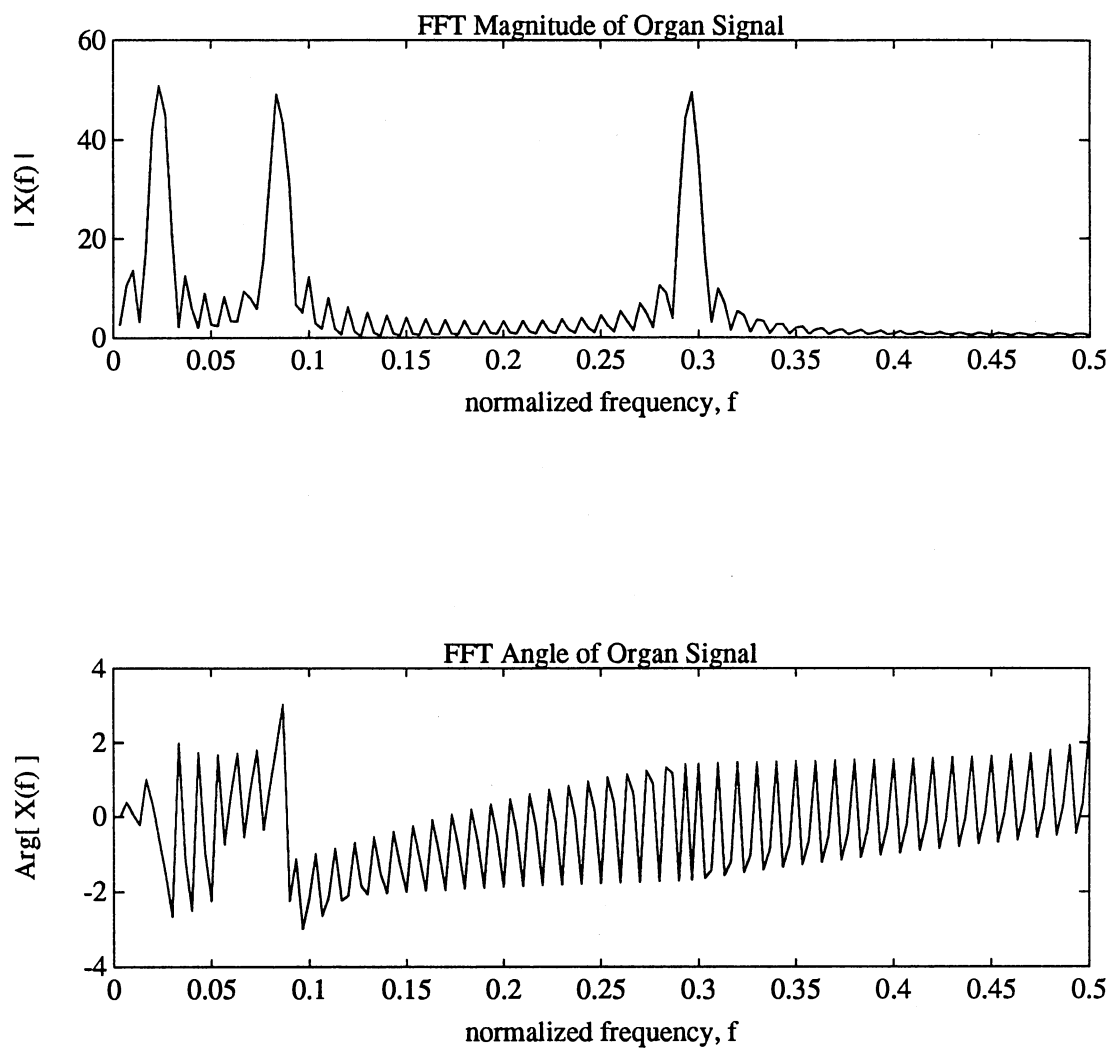


Figure I.11 The Fourier Transformation of an Organ Melody

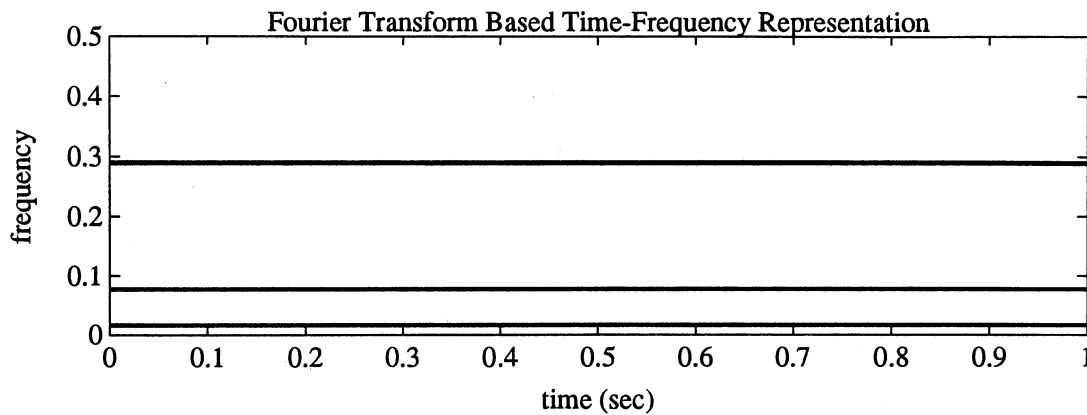


Figure I.12 The Fourier Transformation Based Time-Frequency Representation of the Organ Signal

This simplistic example shows some of the problems faced when designing a signal processing system for time varying signals. The problem becomes more difficult as the signal becomes more complex. For instance, if the organ produced multiple notes at a particular time, then the organ signal is no longer described with one sinusoid but with many sinusoids. More complications result when the organ produces signals that are not strictly sinusoids. These complications are discussed briefly in the next section which discusses speech signals.

Speech Signal Processing

Although the previous example, identifying organ melodies, showed the difficulties involved with processing time varying signals, the actual organ signals were quite simple. To give a more representative signal of what is typically found in signal processing, this

section discusses the processing of speech signals. Many signal processing applications exist for speech signals but recognizing speech, filtering speech, and coding speech are the most common. Recognizing speech is the ultimate goal of many speech processing researchers since it has a wide range of applications and because recognizing speech is a very challenging problem. However, researchers have found that speech signals are complex and, therefore, quite difficult to recognize. Speech coding allows storing and transmitting speech signals efficiently. The more popular coding techniques use models for a speech signal's spectrum. Thus, the organ signal example relates directly to speech coding. In addition to these two speech applications, speech signal filtering is widely implemented using advanced signal processing techniques.

As already mentioned, a speech signal is much more complicated than the simple organ signal. First, a speech signal contains many frequencies at a particular time and, in fact, contains spectral energy from many frequencies over time. Thus, speech signals are not described by one sinusoid at each time but are composed of many sinusoids occurring for several times. Additionally, the sinusoid peak amplitudes are not constant, as was the case for the simple organ signal discussed earlier. Unequal signal amplitudes are evident by the fact that human speech loudness changes continuously. The differences between the simplistic organ signal and a speech signal are exemplified in Figure I.13 which shows how a speech signal decomposes into sinusoids.

The graph in Figure I.13 differs from the corresponding graph in Figure I.3 in two regards. First, each sinusoid contains an extra parameter corresponding to the non-constant peak value of each sinusoid. Second, many sinusoids compose each point. In equation form, a speech signal models as

$$x(t) = \sum_{f=0}^N X(t, f) \sin(2\pi ft + \phi(t, f)), \quad (\text{I.16})$$

where $X(t, f)$ is the Fourier coefficient for a particular frequency and a certain time while

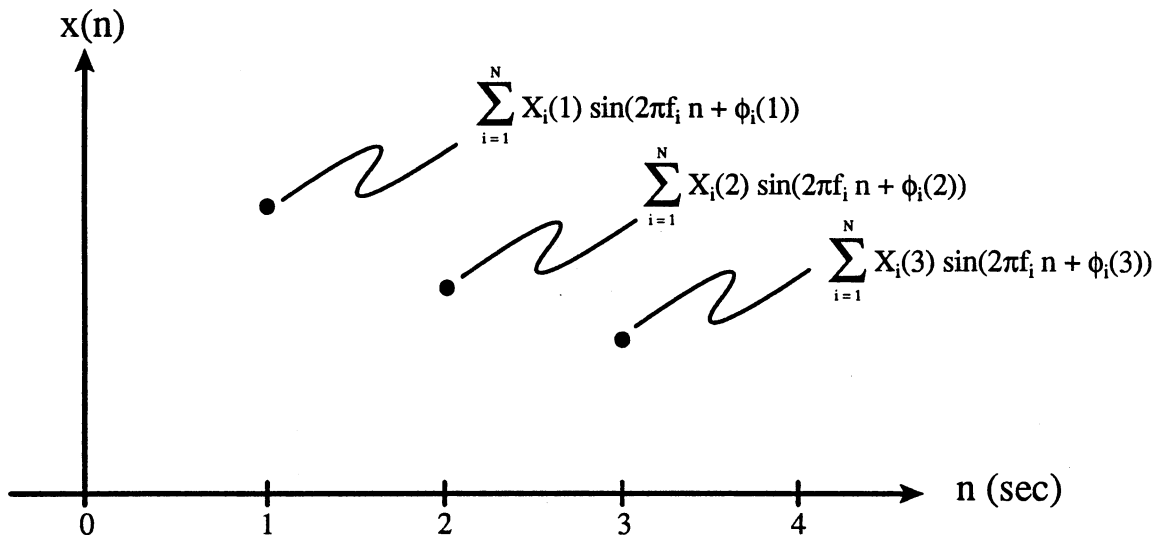


Figure I.13 Speech Signal Sinusoidal Decomposition

$\phi(t, f)$ represents phase for a particular time and frequency. Comparing (I.16) to (I.14) indicates that $4N$ parameters replace the two parameters necessary in (I.14). Thus, speech signals are much more complicated than simple organ signals. However, speech signals do behave nicely in many instances due to naturally occurring properties that are described in the Appendix.

The main question that requires an answer is how signal processing recognizes speech, filters speech, or codes speech. As in identifying organ melodies, the nature of producing speech signals is studied. Additionally, the natural speech reception system (the hearing system) is investigated. Because speech signals are more complicated than organ signals, a discussion of the speech production system and the human hearing system is placed in the Appendix. However, the information in the Appendix can be summarized by stating that the speech production system is nicely described in terms of

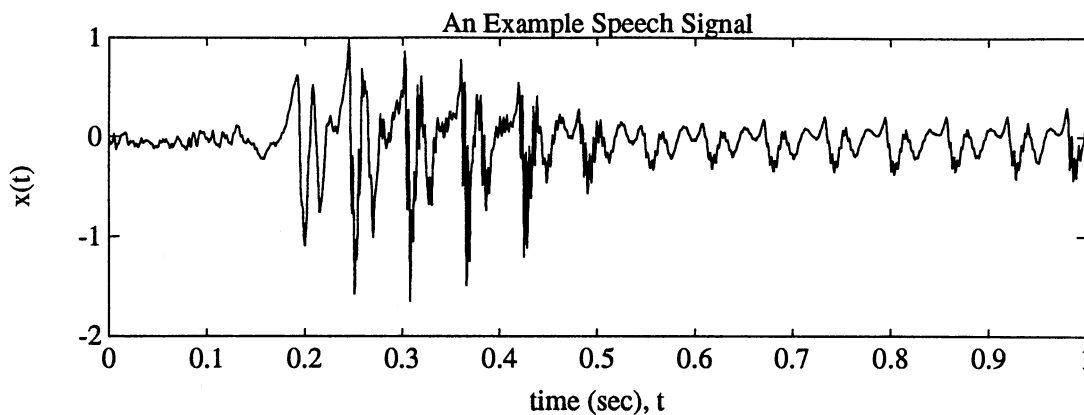


Figure I.14 A Speech Signal Example

sinusoids while the human hearing system transforms sound into a time-frequency representation. Thus, similar to identifying organ melodies, time-frequency representations are used in speech processing.

Illustrating the necessity of time-frequency representations in speech processing is a typical speech signal shown in Figure I.14. The speech signal consists of the first four letters of the word SIGNAL which is described phonetically as /s i g n æ l/. What makes this speech signal interesting is the four distinct groups of sounds. Each distinct group of speech represents a *phoneme*, a linguistic term describing a basic sound in language [Rab78]. As discussed in the Appendix, a particular spectral energy distribution characterizes each phoneme. Thus, the speech signal Fourier transform, shown in Figure I.15, combines four distinct spectral representations. Therefore, the magnitude graph would be identical had the four phonemes been spoken in a different order. Again, the Fourier magnitude signal contains no temporal information.

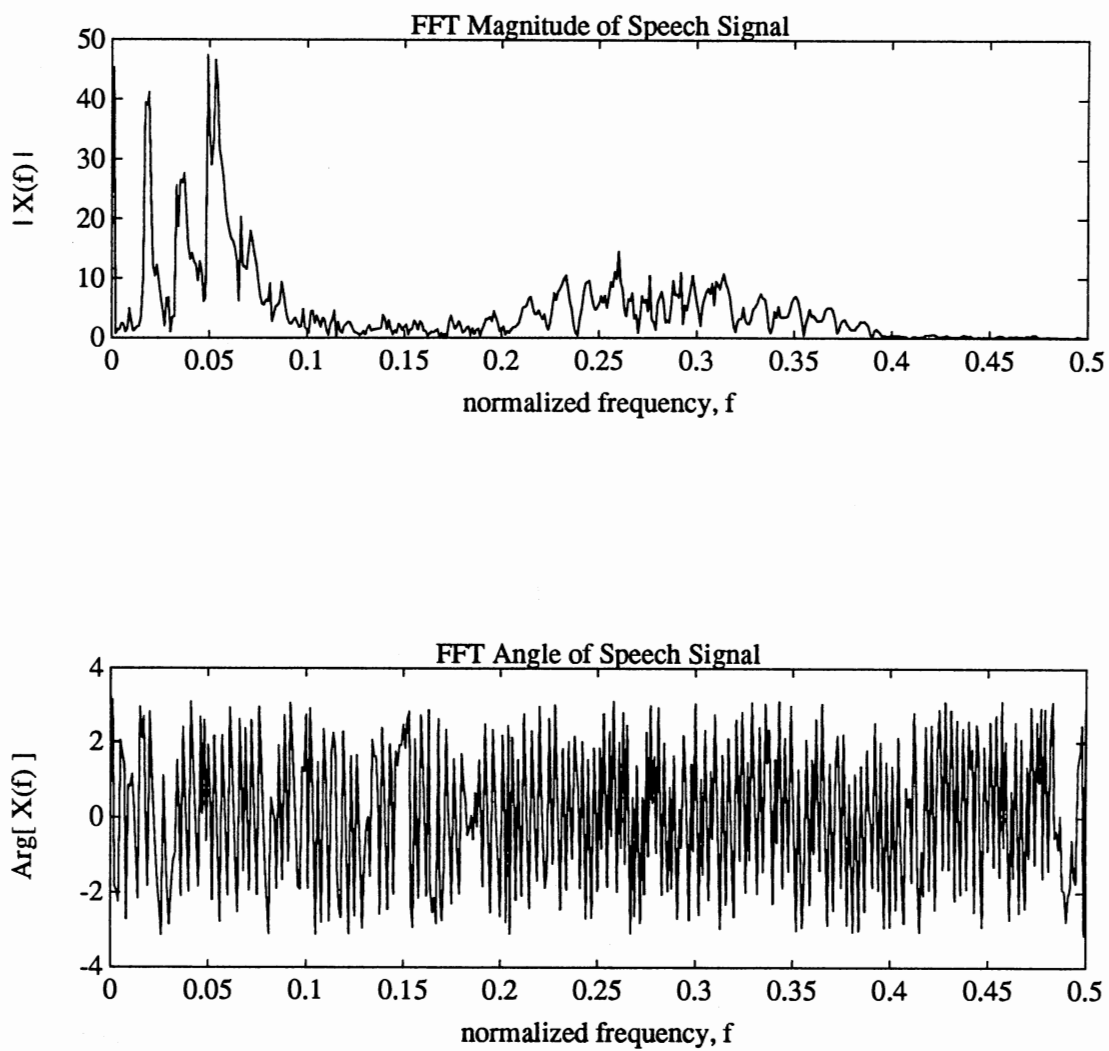


Figure I.15 The Fourier Transformation of the Speech Signal

Similar to organ signal processing, a speech signal time-frequency representation is desirable. For the speech signal, shown in Figure I.14, a time-frequency representation should easily allow determining the distinct sounds. The time-invariant based Fourier transform magnitude obviously failed. Thus, new techniques are required to produce useable time-frequency representations.

Research Goal

This introductory chapter has presented some basic ideas concerning signal processing. First, and foremost, is the abstract signal processing model that describes all signal processing in terms of an input, a system, and an output. Second, if a signal processing system is a linear time-invariant system, then a convolution integral of the input and the impulse response describes the output. Similarly for linear time-invariant systems, the product of the input Fourier transformation and the impulse response Fourier transform magnitudes describes the system output. These two descriptions comprise numerous signal processing techniques, most of which use the Fourier spectrum.

Unfortunately, the world is not perfect and rarely are real world systems linear or time-invariant. Nonlinear system theory is currently under a great amount of investigation and will not be discussed further. Time varying system theory has not been researched as much as nonlinear systems since many signals are, at worst, short-time time-invariant. However, techniques devised to handle short-time time-invariant signals perform poorly in signal regions where the signal rapidly changes (transition regions). Therefore, time varying signal processing theory deserves more research than it has received.

The goal of this thesis is to characterize time varying signals and to use this understanding in developing new signal processing techniques. Particularly, since Fourier transform magnitudes and energies (squared magnitudes) are so important when the system is linear and time-invariant, a time varying version of the Fourier transform mag-

nitute is the ultimate goal. If a time varying Fourier transformation magnitude is discovered, many signal processing applications could benefit. The problem is that time varying signal processing has been researched over the years without significant success. However, the possibility of finding meaningful time varying signal processing techniques is not excluded. Therefore, this research attempts to find anything that can advance signal processing methods of time varying signals.

Research Contributions

This thesis begins investigating time varying spectral energy techniques in Chapter II by thoroughly reviewing the previous work performed in the area. This review indicates that the spectrogram is by far the most widely used time varying spectral energy technique. However, because the spectrogram produces poor spectral estimates during transition regions and in highly nonstationary signals, alternate methods are necessary. Of all the alternate methods discussed, the Wigner distribution, which has been the subject of a considerable amount of research lately, is shown to be quite important for understanding alternate time varying spectral energy techniques. As such, this thesis devotes all of Chapter III to describe the Wigner distribution and to compare the Wigner distribution performance to the spectrogram.

Chapter III provides evidence that, in general, the Wigner distribution is not capable of replacing the spectrogram for estimating time varying spectral energies. However, Chapter III does provide enticing examples that indicate that the Wigner distribution may provide additional time varying spectral energy information when compared to the spectrogram. Thus, two questions remain unanswered. First, does any technique exist for replacing the spectrogram as a spectral energy estimator for time varying signals? Second, does the Wigner distribution or any other time varying spectral energy technique actually provide additional time varying spectral energy information?

Chapter IV addresses the first question by introducing the correlation approach to time-frequency representations. The correlation approach to time-frequency representations (correlation approach, for short) allows studying various techniques for estimating time varying spectral energies in a general setting. The general setting is developed by introducing three new quantities: the time-frequency autocorrelation function, the time-frequency expected value, and the time-frequency spectral energy. The time-frequency spectral energy represents alternate time varying spectral energy techniques and is defined as the Fourier transformation of the time-frequency autocorrelation function with respect to the time difference variable. The time-frequency autocorrelation function, which is similar to a nonstationary autocorrelation function, is defined using a transformation of time variables and the time-frequency expected value. The time-frequency expected value is defined as an expected value that can be computed using arbitrary averages. Chapter IV found that useful averages were the time average, the ensemble average, and the one sample function ensemble average. Using these three averages and various time variable transformations, the correlation approach was able to describe the periodogram, the power spectral density, the Wigner distribution, the Rihaczek distribution, and Page's instantaneous power spectrum. In addition, Chapter IV provided proofs of several general properties that allowed comparing different time-frequency spectral energies, and thus, different time varying spectral energy techniques. As a result of these comparisons, Chapter IV concludes that no time-frequency spectral energy seems capable of replacing the spectrogram for estimating time varying spectral energies.

Chapter V addresses the second question which asked whether the Wigner distribution or other time-frequency spectral energies provide additional time varying spectral energy information compared to the spectrogram. Chapter V attempted to answer this question by investigating the performance of the Wigner distribution on a time varying signal processing application. The application is the nonstationary indicator which depends on time and states whether a signal is stationary or nonstationary. Noting that

Chapter III indicated that the Wigner distribution marginal density and the first order moment might provide time varying spectral energy information, Chapter V extended this idea. The extension involved defining five Wigner distribution features: the instantaneous power, the instantaneous frequency, the Wigner distribution variance, the Wigner distribution skewness, and the Wigner distribution kurtosis. The higher order moments were included because the Wigner distribution is not Gaussian distributed.

Chapter V applied these features to a multilayer perceptron which implemented the nonstationary indicator. Because the multilayer perceptron performs arbitrary mappings, implementing the nonstationary indicator with the multilayer perceptron allowed investigating whether the Wigner distribution features actually contained time varying spectral energy information. In addition to using the multilayer perceptron, a new neural network, the Volterra expansion multilayer perceptron, is discussed and applied to the nonstationary indicator. The Volterra expansion multilayer perceptron introduces a second order Volterra expansion into the multilayer perceptron's hidden layers. Chapter V provides an example that shows a Volterra expansion multilayer perceptron that trains significantly faster than a comparable multilayer perceptron. However, the training times for the nonstationary indicator did not differ significantly between the Volterra expansion multilayer perceptron and the multilayer perceptron.

After training both the multilayer perceptron and the Volterra expansion multilayer perceptron, Chapter V indicates that the Wigner distribution features do not contain enough time varying spectral energy information for the neural networks to generalize the mapping between the Wigner distribution features and the nonstationary indicator output. Therefore, the Wigner distribution features do not directly contain useful time varying spectral energy information in the context of the nonstationary indicator. Because many of the other time-frequency spectral energies contain the instantaneous power and similar other features, Chapter V concludes that time-frequency spectral energies, in general, do not contain enough time-frequency spectral energy information to implement the nonsta-

tionary indicator. The reader should note however that applying additional signal processing techniques to the nonstationary indicator output may allow robust nonstationary indicator implementations. In addition, modifying the nonstationary indicator training process may produce better results.

Chapter VI presents a review of all thesis findings. The overall thesis conclusion is that time-frequency spectral energies do not seem capable of replacing the spectrogram for estimating spectral energies for time varying signals. In addition, time-frequency spectral energies do not contain a sufficient amount of time varying spectral energy information to implement a seemingly easy time varying signal processing application (the nonstationary indicator).

CHAPTER II

REVIEW OF PAST WORK

Introduction

The previous chapter introduced problems associated with time varying signals. Most, if not all, signal processing techniques assume a linear time-invariant system. This assumption presents difficulties when a system varies in time. As such, researchers have investigated different methods of analyzing time varying signals. For the most part, this research defines either a time varying version of Fourier transform energy (spectral energy) or develops a time-frequency representation that contains spectral energy information. Two reasons exist for concentrating on time varying spectral energy techniques. First, most signal processing techniques are based on spectral properties. Second, electrical engineers are comfortable with spectral information. Thus, electrical engineers can easily accept new time varying spectral energy techniques. This chapter discusses most of the time varying spectral energy techniques introduced through the years with a majority of these techniques consisting of time-frequency representations. In addition to discussing the different time varying spectral energy techniques, a comparison between the techniques is also provided.

Before presenting the earlier works in time-frequency representations, two questions require discussion. First, why is the Fourier transformation not usable on time varying signals? Second, what properties are desirable in a time-frequency representation? The two examples discussed in Chapter I indicated why the Fourier transformation is not usable for time varying signals since the Fourier transformation

energy contains no temporal information. Looking at the Fourier transformation pair (I.7) notice that arbitrary signals, $x(t)$, decompose into basis functions, $e^{i2\pi ft}$. Since the basis functions are sinusoids, arbitrary signals decompose into sinusoids of different frequencies, phases, and amplitudes. One problem with decomposing time varying signals in terms of sinusoids is that sinusoids are time-invariant. Decomposing time varying signals with time-invariant basis functions precludes decomposing time varying signals. Another problem with sinusoidal basis functions is that sinusoids are not time-limited. A *time-limited signal* contains non-zero values for only a finite amount of time. Thus, a signal $x(t)$ is time-limited if

$$x(t) = \begin{cases} 0, & \forall t < t_1 \\ 0, & \forall t > t_2 \end{cases} \quad (\text{II.1})$$

where t_1 and t_2 represent finite times and $t_2 > t_1$. Because sinusoidal basis functions are not time-limited, decomposing time-limited signals is inefficient.

Since the Fourier transformation inefficiently decomposes time varying signals, alternate techniques have been developed over the years. These techniques consist of three basic methods. The first method modifies the Fourier transformation so useful spectral energies arise from time varying signals. The second method defines time-frequency representations as distributions of spectral energy and signal power. The second method produces time-frequency representations that mimic two-dimensional probability density functions. The third method replaces the complex exponential basis function with alternate basis functions designed specifically for time varying signals. The new basis functions decompose signals in terms of time and frequency instead of frequency alone. This chapter contains examples extracted from these three techniques.

The second question that requires discussion is what properties time-frequency techniques contain. Because the three time varying spectral energy methods differ widely, the methods each contain different properties and therefore their usefulness

varies. Unfortunately, no technique fulfills all the desired time varying spectral analysis properties. Depending upon the application, some properties are obviously more desirable than others. Since the three spectral energy techniques address different properties, the properties of each are separately discussed.

Because the first method, adapting the Fourier transformation to time varying signals, extends the Fourier transformation energy, alternate techniques should adhere to properties of the Fourier transformation modulus squared. Some of these properties arise from the squared modulus function while the other properties arise from Fourier transformation properties. All the properties will be described in terms of an arbitrary signal, $x(t)$, and an arbitrary time-frequency representation designated, $TF_x(t, f)$. Notice that the time-frequency representation, $TF_x(t, f)$, explicitly depends on both time and frequency. Throughout this thesis, $TF_x(t, f)$ will imply some form of time varying spectral energy technique. The squared modulus function enacts the following properties on $TF_x(t, f)$:

Property II.1: $TF_x(t, f) \in \mathfrak{R}$

Property II.2: $TF_x(t, f) \geq 0.0$.

Thus, techniques that extend the Fourier transformation should produce real and non-negative values.

Many other properties exist by comparing a signal's time domain representation with its time-frequency representation. In the following properties, the \Leftrightarrow symbol implies that the time domain signal property exists if and only if the time-frequency representation property exists. Similarly, the \Rightarrow symbol implies that the time-frequency representation property holds if the time domain signal property holds.

Property II.3: $x(t) = x_1(t) + x_2(t)$

\Leftrightarrow

$$TF_x(t, f) = TF_{x_1}(t, f) + TF_{x_2}(t, f)$$

Property II.4:

$$x(t) = \int_{-\infty}^{\infty} x_1(\xi) \cdot x_2(t - \xi) d\xi$$

$$\Leftrightarrow$$

$$TF_x(t, f) = TF_{x_1}(t, f) \cdot TF_{x_2}(t, f)$$

Property II.5:

$$x(t) = 0, \quad \forall |t| > T$$

$$\Rightarrow$$

$$TF_x(t, f) = 0, \quad \forall |t| > T$$

Property II.6:

$$X(f) = 0, \quad \forall |f| > F$$

$$\Rightarrow$$

$$TF_x(t, f) = 0, \quad \forall |f| > F$$

Property II.7:

$$x(t) = x_1(t) \cdot e^{i2\pi f_1 t} \Leftrightarrow X(f) = X_1(f - f_1)$$

$$\Leftrightarrow$$

$$TF_x(t, f) = TF_{x_1}(t, f - f_1)$$

Property II.8:

$$x(t) = g(TF_x(t, f)).$$

These eight properties considerably constrain $TF_x(t, f)$. Properties II.3 and II.4, linearity and convolution, form the basis of linear system theory and signal processing. Thus, these two properties allow for useful time-frequency representations. Property II.5, finite time support, is important for applications that use time-frequency representations for detecting the presence of signals. Signal processing applications that detect signals typically search for significant energies over time. If no finite time support property exists, then detecting signals is probably not possible. Properties II.6 and II.7, finite frequency support and frequency shift invariance, force time-frequency representation graphs at a particular time to resemble a normal Fourier transformation. Property II.8 implies that

the original signal $x(t)$ can be obtained from its corresponding time-frequency representation, $TF_x(t, f)$, using some function $g(\cdot)$. Property II.8 allows the possibility of coding $x(t)$ in the time-frequency domain since $x(t)$ is recoverable from $TF_x(t, f)$.

These eight spectral energy based properties define an ideal time-frequency representation based upon extending the Fourier transformation. Unfortunately, extending the Fourier transformation to time varying signals nullifies some of these properties.

However, if a property is not applicable, the remaining properties might still allow useful signal processing.

Properties derived from the second method, where time-frequency representations act as two-dimensional probability density functions (pdfs), result directly from probabilistic properties. If $TF_x(t, f)$ represents a two-dimensional pdf, the following properties would hold:

Property II.9: $TF_x(t, f) \in \Re$

Property II.10: $TF_x(t, f) \geq 0$

Property II.11:
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} TF_x(t, f) dt df = 1$$

Property II.12:
$$m_k(t) = \int_{-\infty}^{\infty} \xi^k TF_x(t, \xi) d\xi$$

Property II.13:
$$m_k(f) = \int_{-\infty}^{\infty} \xi^k TF_x(\xi, f) d\xi$$

Property II.14:
$$\text{Prob}[TF_x(t, f) > 0, \forall t_1 \leq t \leq t_2, f_1 \leq f \leq f_2] = \int_{t_1}^{t_2} \int_{f_1}^{f_2} TF_x(t, f) dt df.$$

Notice that the first Properties II.9 and II.10 correspond directly to the Fourier transformation energy based properties. Thus, these two properties are quite desirable. Property II.11 requires the total probability to equal 1. Properties II.12 and II.13 allow computing

moments on the time-frequency representation $TF_x(t, f)$. Note that two moments in Properties II.12 and II.13 are functions of time and frequency, respectively. Although not explicitly stated, these two properties imply that the moments contain relevant signal information that is similar to probability moments such as marginal distributions, means, and variances. Property II.14 allows computing the amount of signal energy for an area defined by time and frequency. If a time-frequency representation contains these six properties, then analysis techniques from probability theory directly apply. Applying probability theory to time-frequency representations allows most engineers to interpret time-frequency representations easily.

The third method involves defining new basis functions to replace the complex exponentials of the Fourier transformation. Let

$$x(t) = \int_{-\infty}^{\infty} X(t, f) \phi(t, f) df, \quad (\text{II.2})$$

where $\phi(t, f)$ are the basis functions and $X(t, f)$ correspond to the decomposing coefficients. The two main properties a basis function should possess are:

Property II.15:
$$\int_{-\infty}^{\infty} \phi(t_1, f) \phi(t_2, f) df = \begin{cases} 0, & \forall t_1 \neq t_2 \\ c, & \forall t_1 = t_2 \end{cases}$$

Property II.16: with $c = 1$.

Property II.15 states that basis functions are orthogonal to each other. Orthogonality implies that each basis function represents a unique portion of the signal. Also, orthogonality implies that Parseval's relation

$$\int_{-\infty}^{\infty} |X(t, f)|^2 df = \int_{-\infty}^{\infty} |x(t)|^2 dt \quad (\text{II.3})$$

holds; and thus, both representations contain the same amount of energy. Property II.16 indicates that the orthogonal basis set is orthonormal. The goal of using new basis functions is to more efficiently represent time varying signals.

Obtaining the Properties II.1 through II.16 will not be possible with a single technique. However, these properties guide designers of time-frequency representations. The remaining sections of this chapter give the important techniques previously researched. The spectrogram, the most widely used time-frequency representation, is discussed in detail since it intuitively extends the Fourier transformation and is, thus, understandable. The remaining time-frequency representations are covered only in enough detail to see the relevance of each representation.

Previous Time-Frequency Representations

The Spectrogram

The spectrogram was probably the first time varying spectral analysis technique to obtain wide support in signal processing. The spectrogram achieved wide support by extending the basic Fourier transformation. A problem in using the Fourier transformation on time varying signals is that (I.7) and (I.8), the Fourier transformation pair, requires signal knowledge for all time. To alleviate this problem, the limits of integration in (I.7) are changed such that

$$X(t, f) = \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} x(\xi) e^{-2\pi f \xi} d\xi, \quad (\text{II.4})$$

where $x(\xi)$ is assumed time-invariant for $\xi \in [t - T/2, t + T/2]$. Equation (II.4) formulates the Fourier transformation into a time-frequency representation whose spectral estimates

are computed locally on time-limited signals (although sometimes referred to as a windowed transform where the transform is computed on a short term basis, this viewpoint is discussed later). Spectral energy is computed as

$$S_x(t, f) = \left| \int_{t-T/2}^{t+T/2} x(\xi) e^{-i2\pi f \xi} d\xi \right|^2, \quad (\text{II.5})$$

which is the squared modulus function of (II.4). Both (II.4) and (II.5) assume that the signal, $x(t)$, varies globally but not locally. Recall that global, but not local, variation occurred for most of the organ signals and speech signals described in Chapter I.

The time-limited Fourier transformation, (II.5), can be rewritten as

$$S_x(t, f) = \int_{-\infty}^{\infty} h(t - \xi) x(\xi) e^{-i2\pi f \xi} d\xi \quad (\text{II.6})$$

where

$$h(\xi) = \begin{cases} 1, & \forall \xi \in [t - T/2, t + T/2] \\ 0, & \text{o.w.} \end{cases} \quad (\text{II.7})$$

If $h(\xi)$ changes, the time-limited Fourier transformation is called the *short-time Fourier transformation* and is defined as

$$X(t, f) = \int_{-\infty}^{\infty} x(\xi) h(\xi - t) e^{-i2\pi f \xi} d\xi. \quad (\text{II.8})$$

Therefore,

$$S_x(t, f) = |X(t, f)|^2 = \left| \int_{-\infty}^{\infty} x(\xi) h(\xi - t) e^{-i2\pi f \xi} d\xi \right|^2 \quad (\text{II.9})$$

defines the spectral energy [Pap62], or the *spectrogram*, where $h(\xi)$ represents any window function. Thus, the spectrogram modifies the Fourier transformation energy by limiting the number of signal values that are non-zero.

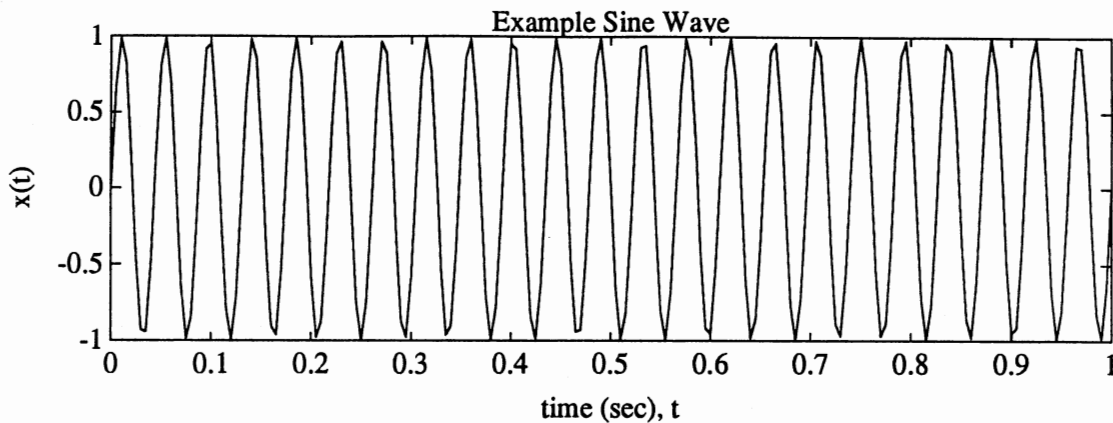


Figure II.1 A Time-Invariant Sine Wave

An example should clarify the differences between a normal Fourier transformation and a short-time Fourier transformation. Figure II.1 displays the graph of a sinusoid given by

$$s(t) = \sin(2\pi t f), \quad (\text{II.10})$$

where $f = 23$ hertz. Suppose the spectral estimate of the signal at time $t = 0.5$ sec is desired. Figure II.2 displays the graph of the rectangular window which is defined by

$$h(\xi) = \begin{cases} 1, & \forall \xi \in [t - T/2, t + T/2] \\ 0, & \text{otherwise} \end{cases}. \quad (\text{II.11})$$

with $T=0.25$ seconds. Applying the rectangular window to the sine wave produces the signal also given in Figure II.2. Figure II.3 shows the graph for the spectral magnitudes of both the sine wave and the rectangular windowed sine wave. Notice how the rectangular window reduces the resolution of the normal Fourier transformation, adds noise to the remaining spectra, and creates spectral leakage. Windows may reduce resolution and add noise, but the short-time Fourier transformation performs better on time varying signals.

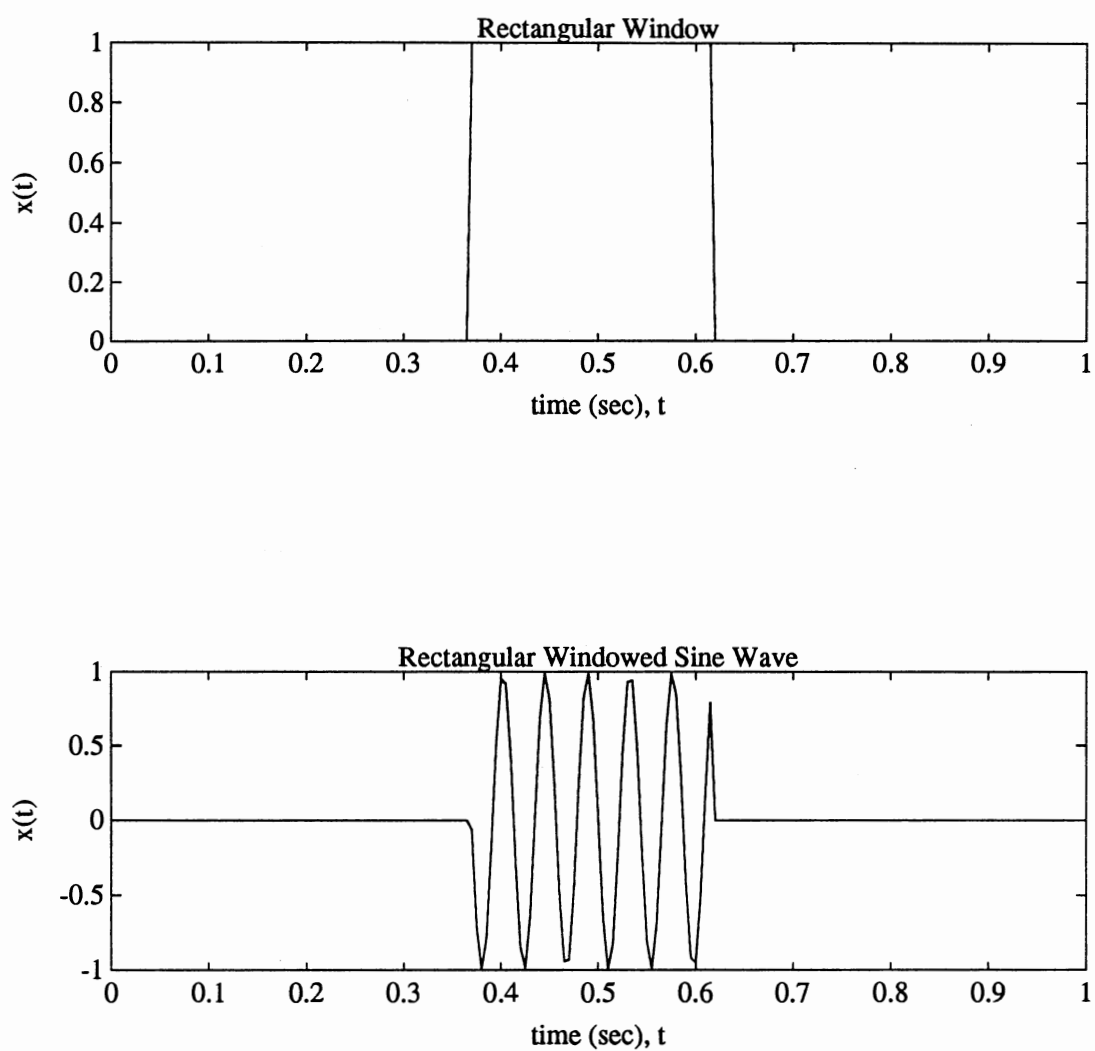


Figure II.2 A Rectangular Window and the Rectangular Windowed Sine Wave

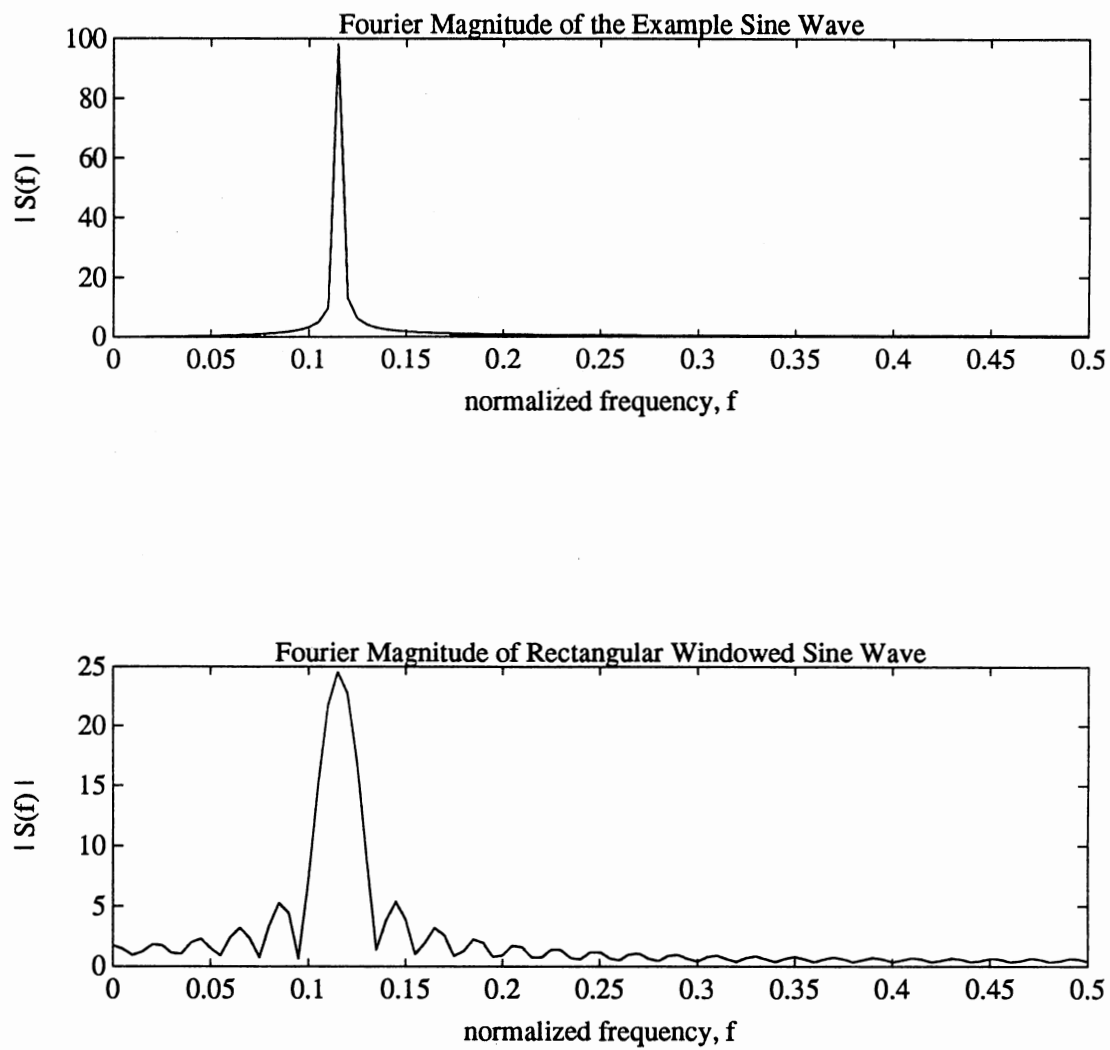


Figure II.3 The Fourier Transform Magnitude of the Sine Wave and Rectangular Windowed Sine Wave

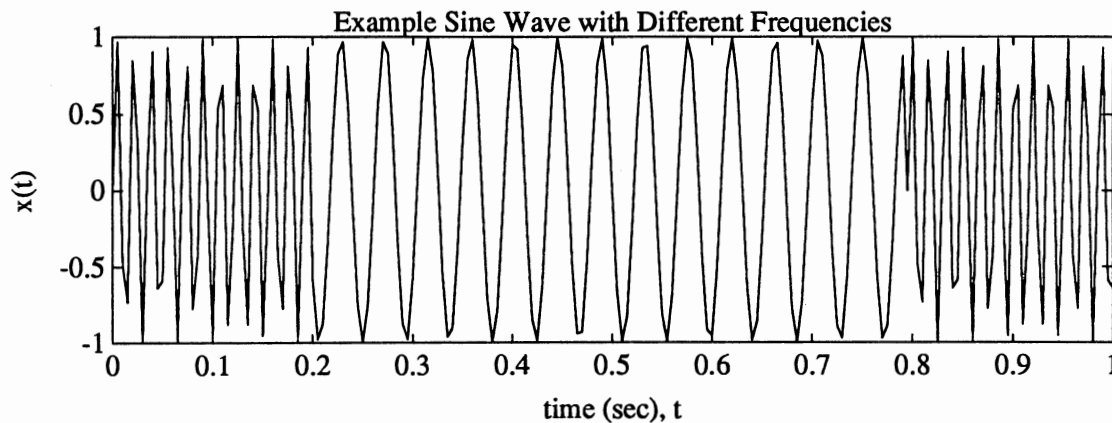


Figure II.4 A Time Varying Sine Wave

Figure II.4 displays the graph of a time varying sine wave which is defined by (II.10) but with frequencies

$$f = \begin{cases} 58\text{Hz} & \forall t \in [0.0, 0.2] \\ 23\text{Hz} & \forall t \in (0.2, 0.8) \\ 58\text{Hz} & \forall t \in [0.8, 1.0] \end{cases} \quad (\text{II.12})$$

This signal (II.12) is the same as (II.10) for times between 0.2 and 0.8 seconds. Figure II.5 displays the graph of the Fourier transformation magnitude of the time varying sine wave along with the rectangular windowed time varying sine wave. Notice that the rectangular windowed time wave spectrum at time 0.5 seconds is identical to the one computed for the time-invariant sine wave. Further notice that the normal Fourier transformation magnitude at time 0.5 seconds differs from Figure II.3. Because the Fourier transformation estimates the presence of a 23 Hz and 58 Hz signal at time 0.5 seconds

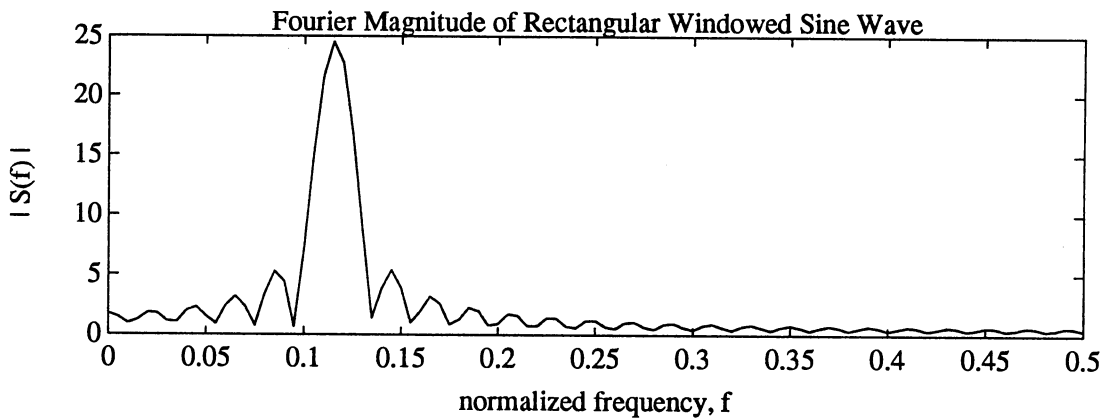
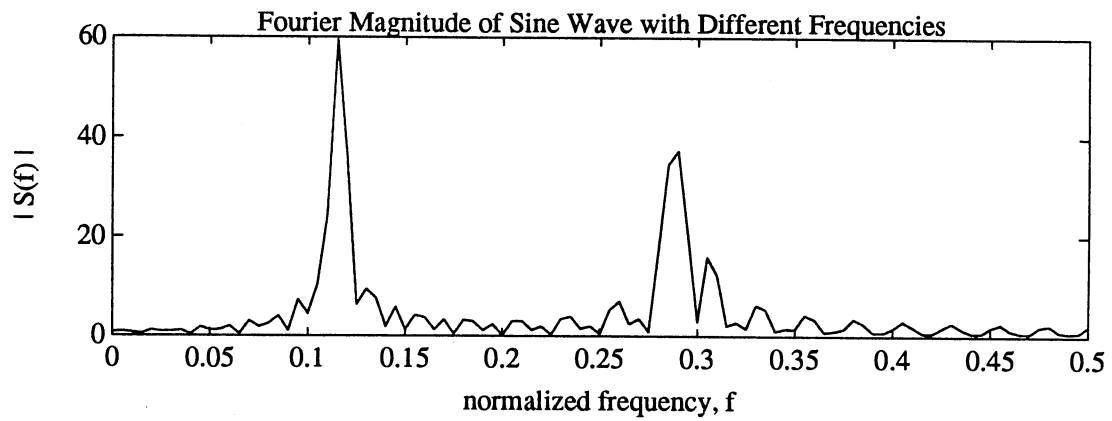


Figure II.5 The Fourier Transform Magnitude of the Time Varying Sine Wave and Rectangular Windowed Time Varying Sine Wave

and because the short-time Fourier transformation estimates only the 23 Hz signal, the short-time Fourier transformation produces better spectral estimates. Thus, although the rectangular windowed short-time Fourier transformation estimate induces noise and resolves less frequency for time varying signals, the short-time Fourier transformation is superior to the Fourier transformation for such cases.

A question arises whether any methods exist that reduce noise and resolve more spectral detail. If both noise reduction and more frequency resolution are necessary, no method exists. However, replacing the rectangular window with other windows reduces noise but resolves less frequency. Many different windows exist but the most popular, at least in speech processing, is the Hamming window defined by

$$h(t) = .54 + .46 \cos\left(\frac{2\pi t}{L}\right), \quad (\text{II.13})$$

where L is the window length. Another window used mainly for theoretical work is the Gaussian window which is defined as

$$h(t) = e^{-\frac{\sigma^2 t^2}{2(n-1)^2}}, \quad (\text{II.14})$$

where σ determines the window width for $t \in [-L, L]$. The graphs in Figure II.6 show the differences in time shifted versions of the rectangular, Hamming, and Gaussian windows. Figure II.7 indicates the effect each window has on the sine wave defined in (II.12). Figure II.8 displays the graphs for the Fourier magnitudes of the windowed signals. The spectral graphs for the different windowed sine waves indicate that the Hamming and Gaussian windows reduce noise at a cost of a wider mainlobe. The Hamming window resolves more frequency than the Gaussian window. Thus, the Hamming window is the best window and the Gaussian window is the next best window. The reader should note that many other windows exist (Kaiser, Bartlett, etc.) and each window is useful for many applications. In fact, a window is typically chosen based on time-bandwidth products [Ope78] but the rectangular, Hamming, and Gaussian windows

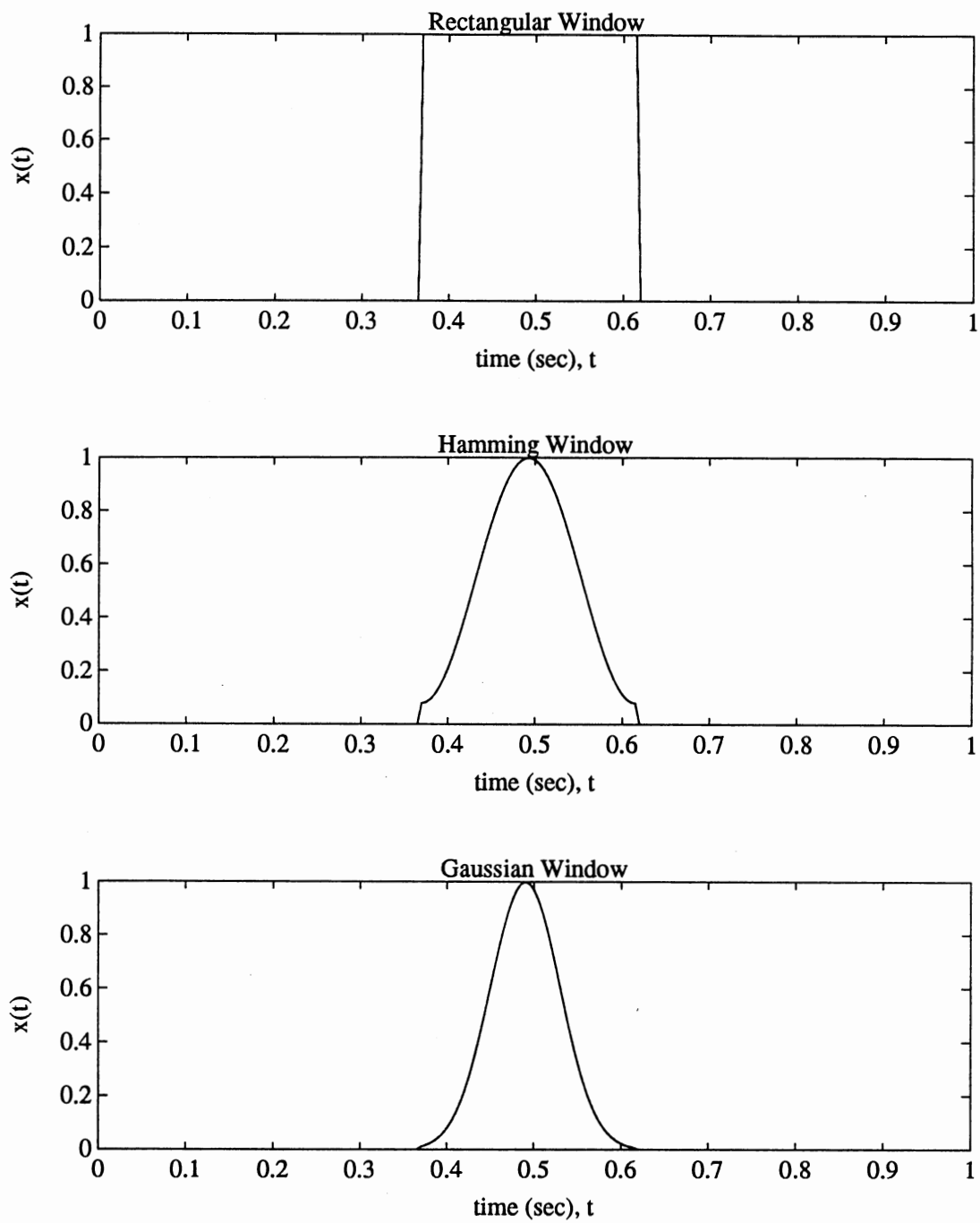


Figure II.6 Examples of a Rectangular, Hamming, and Gaussian Window

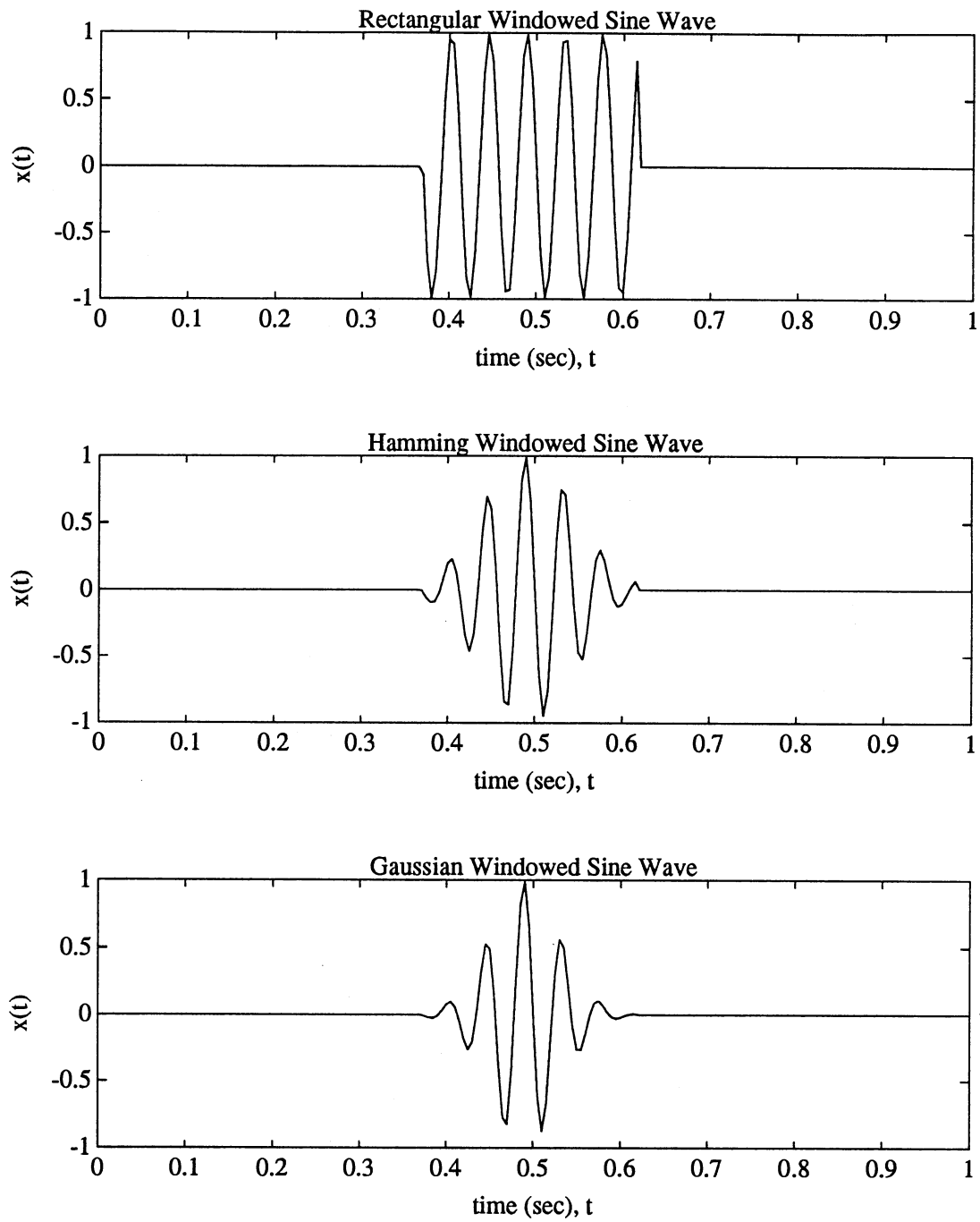


Figure II.7 Different Windows Applied to the Sine Wave

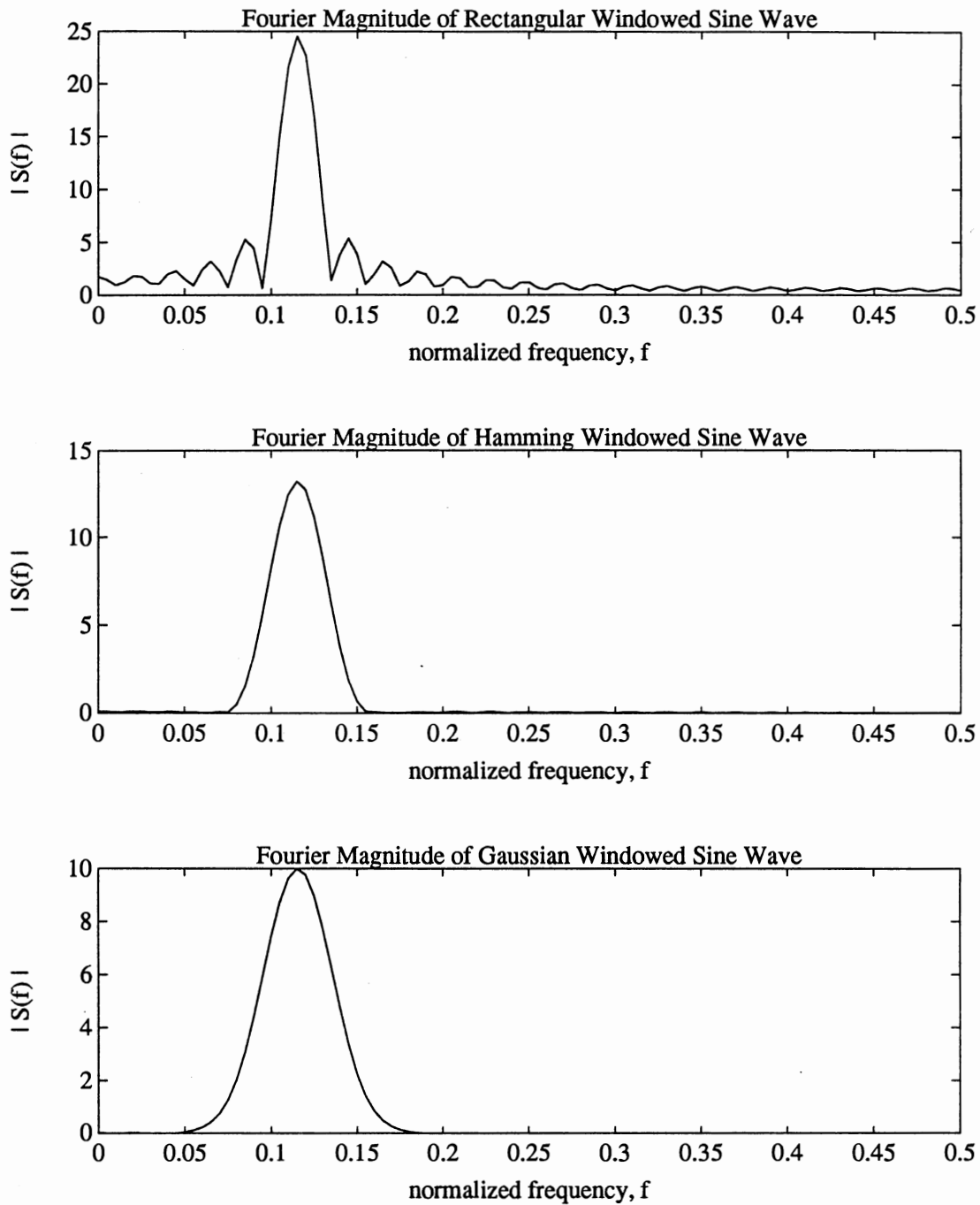


Figure II.8 Fourier Transformation Magnitude of the Rectangular, Hamming, and Gaussian Windowed Sine Wave

were evaluated to give the reader an idea of how windows differ. However, all remaining windowing operations in this thesis use the Gaussian window since it performs similar to the Hamming window and allows simpler theoretical evaluations.

Why did the rectangular window perform so poorly? Analyzing the Fourier transformation of the window functions answers this question. Because multiplying a window and signal together in the time domain corresponds to convolving the window's Fourier transformation and the signal's Fourier transformation, the window's Fourier transformation should ideally equal an impulse function. In finite time, a windowed Fourier transformation cannot be an impulse function so windows are designed to produce impulsive shaped spectrums. This discussion leads into the idea a time-bandwidth trade-off where the time resolution multiplied by the frequency resolution equals a constant. This being the case, an infinitely sharp frequency resolution (an impulse spectrum) is impossible except for infinitely long windows.

Figure II.9 shows the magnitude graphs of the three windows. Notice that the rectangular window is more impulsive for the mainlobe but contains large sidelobes. The Hamming and Gaussian windows, however, contain reduced sidelobes at a cost of wider mainlobes. Also notice that the Gaussian window mainlobe is larger than the Hamming window mainlobe. Thus, the Hamming window best approximates an impulsive function. Observe that during the window examples, as resolution increases, noise increases, and visa versa. Gabor showed that a signal cannot be perfectly resolved in both time and frequency; in fact, Gabor showed that

$$\Delta t \Delta f \geq \frac{1}{2}, \quad (\text{II.15})$$

where Δt represents time resolution and Δf represents frequency resolution [Gab46]. The Fourier uncertainty principle, (II.15), is similar to the Heisenberg uncertainty principle of physics that disallows knowing an electron's exact position and momentum simultaneously. The uncertainty principle constrains the amount of time varying signal

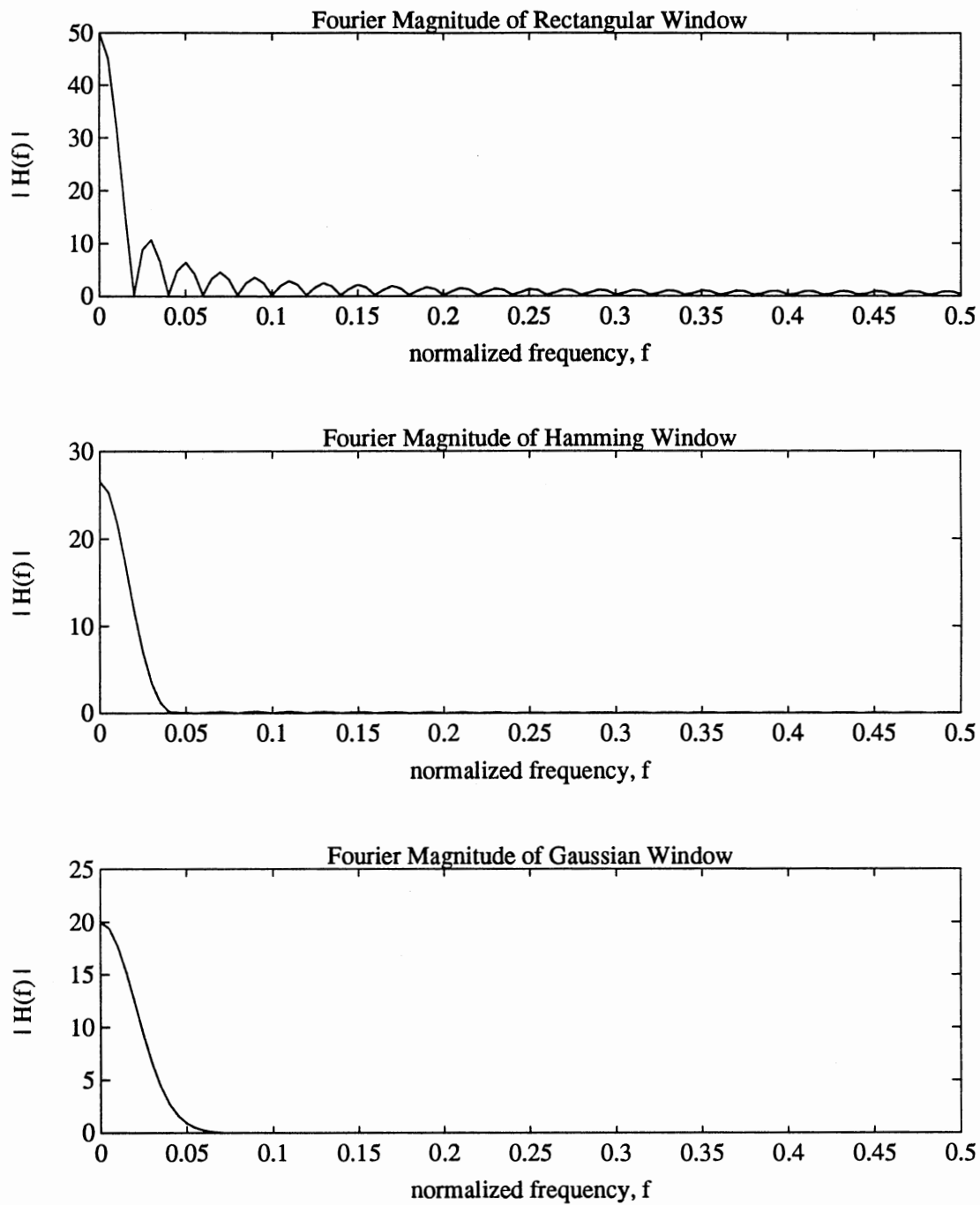


Figure II.9 Fourier Transformation Magnitude of the Rectangular, Hamming, and Gaussian Windows

information available because short signal lengths, which are better for time varying signals, produce coarse spectral estimates. Yet, the more useful time spectral estimates require long signal lengths. Thus, alternate time varying spectral energy techniques attempt to get around (II.15) by manipulating the signal and frequency variables in a time-frequency transformation. This chapter presents some time-frequency representations of this nature.

Although this section points out that the spectrogram does not perform well, signal processors often implement the spectrogram. The spectrogram was originally implemented in 1946 in an apparatus called a sonogram [Koe46]. The sonogram used a bank of non-overlapping bandpass filters, whose response simulated the window functions previously discussed, to plot spectral energy over time [Ste46]. Because the spectrogram is easy to implement, the spectrogram remains a popular time varying spectral energy estimator. The following sections give alternate methods for estimating time varying spectral energies. Before discussing the alternate methods, some terminology must be straightened out. This thesis is investigating methods for estimating time varying spectral energies. The word energy refers to the fact that the Fourier transformation squared magnitude is computed. However, in some of the following discussion, the original authors often used the term power instead of the word energy. The reason for the different wording is that many signals contain infinite energy signals (such as a infinitely long sine wave) but contain finite power signals. Thus, a predicament exists as to what terminology should be used in this thesis. For historical reasons, already established techniques will retain their names regardless if the term power is used. For any new techniques developed by this thesis and for any general discussions concerning time varying spectral energies, the word energy will be used. Although this terminology is sort of confusing, for convenience think of power as energy. One more confusing terminology problem must be addressed. The squared magnitude Fourier transformation of a signal $x(t)$ is denoted $|X(f)|^2$ and is called the Fourier transformation energy density. The term density

is used because integrating $|X(f)|^2$ for all f produces the total energy. The term density is also used in the same respect for the term power. A problem exists for the historical procedures because the term density is often assumed and not used. Thus, in order to reduce confusion, the term density will not be used at all in this thesis. Thus, $|X(f)|^2$ will be referred to as the Fourier transformation energy after Papoulis [Pap62].

Page's Instantaneous Power Spectrum

In 1952, Page [Pag52] described a time varying spectral energy technique based upon Parseval's energy theorem. Additionally, Page constrained signal knowledge so no future signal information was required. Eliminating future signal values allowed real time implementations of Page's technique. Page's technique begins by defining the signal $x(t)$, $\forall t \in (-\infty, T]$, for arbitrary time value T . Additionally, Page's technique uses Parseval's theorem which states

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df \quad (\text{II.16})$$

where

$$X(f) = \int_{-\infty}^{\infty} x(\xi) e^{-i2\pi f \xi} d\xi. \quad (\text{II.17})$$

The value $|x(t)|^2$ is called the *instantaneous power* while $|X(f)|^2$ represents the spectral energy. Parseval's theorem implies that the total signal power equals the total spectral energy.

Page described the signal energy distributed in time and frequency by an energy density function, $\rho(t, f)$. As such, the total signal energy up to time T becomes

$$E_x(T) = \int_{-\infty}^T \int_{-\infty}^{\infty} \rho(t, f) df dt. \quad (\text{II.18})$$

The instantaneous power is defined as the rate of change in energy and obviously it has to be positive. Thus,

$$|x(T)|^2 = \frac{\partial}{\partial T} E_x(T) = \frac{\partial}{\partial T} \int_{-\infty}^T \int_{-\infty}^{\infty} \rho(t, f) df dt. \quad (\text{II.19})$$

Applying Leibnitz's rule [Buc78],

$$\frac{\partial}{\partial t} \int_{y(t)}^{x(t)} h(t, \xi) d\xi = \frac{\partial x(t)}{\partial t} h(t, x(t)) - \frac{\partial y(t)}{\partial t} h(t, y(t)) + \int_{y(t)}^{x(t)} \frac{\partial h(t, \xi)}{\partial t} d\xi, \quad (\text{II.20})$$

to (II.20) yields

$$\begin{aligned} |x(T)|^2 &= \frac{\partial T}{\partial T} \int_{-\infty}^{\infty} \rho(T, f) df - 0 + 0 \\ &= \int_{-\infty}^{\infty} \rho(T, f) df. \end{aligned} \quad (\text{II.21})$$

Therefore, for arbitrary time T , $\rho(t, f)$ was termed an *instantaneous power spectrum* because $\rho(t, f)$ contains the instantaneous power. Note that an instantaneous power spectrum does not equal a spectral energy but is proportional to the spectral energy derivative with respect to time. If $\rho(t, f)$ does not equal the spectral energy then why consider using the instantaneous power spectrum? The possibility exists to extract spectral information from $\rho(t, f)$ and use the information in signal processing algorithms. However, (II.21) does not indicate a method for computing $\rho(t, f)$. Developing a method for computing $\rho(t, f)$ begins by letting

$$x_T(t) = \begin{cases} x(t), & \forall t \leq T \\ 0, & \forall t > T \end{cases}. \quad (\text{II.22})$$

Combining (II.16), (II.18) and (II.21) then produces

$$E_x = \int_{-\infty}^T \int_{-\infty}^{\infty} \rho(t, f) df dt = \int_{-\infty}^T |x(t)|^2 dt = \int_{-\infty}^{\infty} |x_T(t)|^2 dt. \quad (\text{II.23})$$

Because (II.16) holds, $\rho(t, f)$ should allow

$$E_x = \int_{-\infty}^{\infty} |X(f)|^2 df = \int_{-\infty}^T \int_{-\infty}^{\infty} \rho(t, f) df dt, \quad (\text{II.24})$$

which implies $|X(f)| = \int_{-\infty}^T \rho(t, f) dt.$ (II.25)

Differentiating (II.25) with respect to T yields

$$\begin{aligned} \frac{\partial}{\partial T} |X(f)|^2 &= \frac{\partial}{\partial T} \int_{-\infty}^T \rho(t, f) dt \\ &= \frac{\partial T}{\partial T} \rho(T, f) - 0 + 0 \\ &= \rho(t, f). \end{aligned} \quad (\text{II.26})$$

Next, $X(f)X^*(f)$ is substituted for $|X(f)|^2$. Applying the inverse Fourier transformation to $|X(f)|^2$ and computing the inverse Fourier transformation's derivative with respect to t produces

$$\rho(t, f) = \frac{\partial}{\partial t} \left\{ \int_{-\infty}^t x(\xi) e^{-i2\pi f \xi} d\xi \int_{-\infty}^t x(\alpha) e^{i2\pi f \alpha} d\alpha \right\}. \quad (\text{II.27})$$

Applying the chain rule to (II.27) yields

$$\begin{aligned} \rho(t, f) &= x(t) \left\{ e^{-i2\pi f t} \int_{-\infty}^t x(t) e^{i2\pi f y} dy + e^{i2\pi f t} \int_{-\infty}^t x(\alpha) e^{-i2\pi f \alpha} d\alpha \right\} \\ &= 2x(t) \text{Re} \{ X_t(f) e^{i2\pi f t} \}, \end{aligned} \quad (\text{II.28})$$

where $X_t(f) = \int_{-\infty}^{\infty} x_T(t) e^{-i2\pi f t} dt.$ (II.29)

Because $x(t)$ is real valued, then $\rho(t, f)$ becomes

$$\rho(t, f) = 2 \text{Re} \{ x(t) X_t(f) e^{i2\pi f t} \}. \quad (\text{II.30})$$

Thus, (II.28) or (II.30) give a method of computing the instantaneous power spectrum, $\rho(t, f)$, for any signal type. Page's approach was interesting because, by using spectral properties, he produced a technique similar to a two-dimensional probability density function. The reference of $\rho(t, f)$ to a two-dimensional probability density function arises since the two marginal distributions,

$$\int_{-\infty}^{\infty} \rho(t, f) df = |x(t)|^2 \quad (\text{II.31})$$

$$\int_{-\infty}^{\infty} \rho(t, f) dt = |X(f)|^2, \quad (\text{II.32})$$

exist and contain energy information. However, informative marginals were expected since Page constrained $\rho(t, f)$ with (II.25).

Although Page's technique produced a time-frequency representation that distributes signal energy in both time and frequency, the instantaneous power spectrum cannot replace the spectrogram because at time t_0 $\rho(t_0, f)$ indicates changes in the spectral energy and not the spectral energy directly. However, Page's instantaneous power spectrum contains time varying signal information in the form of changing spectral energies so the instantaneous power spectrum is useful. The remainder of this chapter discusses other time-frequency techniques that attempt to describe signals in terms of time and frequency.

Turner's Instantaneous Power Spectrum

In 1954, Turner [Tur54] showed two things. First, Turner proved that Page's instantaneous power spectrum, $\rho(t, f)$, is not unique. Second, Turner defined the running autocorrelation function and showed its relevance to the instantaneous power spectrum.

These two items remind us that the instantaneous power spectrum, $\rho(t, f)$, has many degrees of freedom remaining and that powerful relationships exist between the Fourier transformation energy and the autocorrelation function.

Showing non-uniqueness of the instantaneous power spectrum begins by letting

$$\begin{aligned}\rho'(t, f) &= \rho(t, f) + \rho_c(t, f) \\ &= \frac{\partial}{\partial T} |x_T(f)|^2,\end{aligned}\tag{II.33}$$

where

$$\int_{-\infty}^{\infty} \rho_c(t, f) df = 0.\tag{II.34}$$

Substituting (II.34) into (II.33) and using (II.31) yields

$$\begin{aligned}\int_{-\infty}^{\infty} \rho'(t, f) df &= \int_{-\infty}^{\infty} (\rho(t, f) + \rho_c(t, f)) df \\ &= \int_{-\infty}^{\infty} \rho(t, f) df + \int_{-\infty}^{\infty} \rho_c(t, f) df \\ &= \int_{-\infty}^{\infty} \rho(t, f) df = |x(t)|^2.\end{aligned}\tag{II.35}$$

Therefore, even though $\rho'(t, f) \neq \rho(t, f)$ for some t and f , both $\rho'(t, f)$ and $\rho(t, f)$ are valid instantaneous power spectra. Thus, the instantaneous power spectrum is not unique.

Turner's next contribution related a running autocorrelation and an instantaneous power spectrum. Before progressing, background information concerning the regular autocorrelation function and Fourier transformation energy is presented. Noting that ergodic random processes allow replacing ensemble averages with time averages, assuming an ergodic random process $\{x_k(t)\}$ implies that the stationary autocorrelation function defined by

$$\begin{aligned}
R_x(\tau) &= \overset{\Delta}{E}[x(t)x(t+\tau)] \\
&= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^N x_k(t)x_k(t+\tau),
\end{aligned} \tag{II.36}$$

can be replaced by the time average,

$$\begin{aligned}
R_x(\tau) &= \langle x(t)x(t+\tau) \rangle \\
&= \int_{-\infty}^{\infty} x(t)x(t+\tau)dt.
\end{aligned} \tag{II.37}$$

Replacing ensemble averages with time averages allows estimating $R_x(\tau)$ with one sample function, say $x(t)$. Thus, $x(t)$ allows computing the autocorrelation function. Additionally, the Wiener-Knitchine relation allows estimating spectral energy using

$$S_x(f) = \int_{-\infty}^{\infty} R_x(\tau)e^{-i2\pi f\tau}d\tau, \tag{II.38}$$

where $S_x(f)$, called the *power spectral density*. Note the word power is once again used instead of the term energy. The reason for considering the Fourier transformation of the autocorrelation function as a power signal results from the fact that sample functions are not square integrable. As such, the energy would be infinite. Thus the reader can regard the power spectral density as providing "energy" like information, and so the Fourier transformation of the autocorrelation function produces spectral energy estimates.

Turner defined a *running autocorrelation function* for a signal,

$$x_T(t) = \begin{cases} x(t), & \forall t < T \\ 0, & \forall t > T \end{cases}, \tag{II.39}$$

by

$$\begin{aligned}
A(T, \tau) &= \int_{-\infty}^{\infty} x_T(t)x_T(t+\tau)d\tau \\
&= \int_{-\infty}^T x(\xi)x(\xi+\tau)d\xi.
\end{aligned} \tag{II.40}$$

Note that $A(T, \tau)$ is an even function of τ since

$$\begin{aligned} A(T, -\tau) &= \int_{-\infty}^{\infty} x_T(t) x_T(t - \tau) d\tau \\ &= \int_{-\infty}^{\infty} x_T(t + \xi) x_T(\xi) d\xi = A(T, \xi). \end{aligned} \quad (\text{II.41})$$

The running autocorrelation function relates to the instantaneous power spectrum by first applying Leibnitz's rule to (II.41) yielding

$$\frac{\partial}{\partial T} A(T, \tau) = \begin{cases} x(T) x(T + \tau) & \text{if } \tau < 0 \\ x(T) x(T - \tau) & \text{if } \tau > 0 \end{cases}. \quad (\text{II.42})$$

Using the fact that

$$\begin{aligned} |X_T(f)|^2 &= X_T(f) X_T^*(f) = \int_{-\infty}^{\infty} x(\xi) e^{-i2\pi\xi f} d\xi \int_{-\infty}^{\infty} x^*(t) e^{i2\pi t f} dt \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_T(\xi) x_T^*(t) e^{-i2\pi(t - \xi)f} dt d\xi \end{aligned} \quad (\text{II.43})$$

and substituting $\tau = t - \xi$, then

$$\begin{aligned} |X_T(f)|^2 &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} x_T(t) x_T(t - \tau) dt \right\} e^{-i2\pi f \tau} d\tau \\ &= \int_{-\infty}^{\infty} A(T, -\tau) e^{-i2\pi f \tau} d\tau. \end{aligned} \quad (\text{II.44})$$

Rewriting (II.26) in terms of the running autocorrelation produces

$$\begin{aligned} \rho(T, f) &= \frac{\partial}{\partial T} |X_T(f)|^2 = \frac{\partial}{\partial T} \int_{-\infty}^{\infty} A(T, \tau) e^{-i2\pi f \tau} d\tau \\ &= \int_{-\infty}^{\infty} \frac{\partial}{\partial T} A(T, \tau) e^{-i2\pi f \tau} d\tau. \end{aligned} \quad (\text{II.45})$$

Hence, the instantaneous power spectrum (a spectral energy rate of change) equals the Fourier transformation of the running autocorrelation function rate of change. The relationship between the instantaneous power spectrum and the running autocorrelation function is analogous to the Wiener-Knitchine relation.

Substituting the previous running autocorrelation function rate of change equation (II.42) into (II.45) yields

$$\begin{aligned}\rho(t, f) &= \int_0^{\infty} x(t)x(t-\tau)e^{-i2\pi f\tau}d\tau + \int_{-\infty}^0 x(t)x(t+\tau)e^{-i2\pi f\tau}d\tau \\ &= \int_{-\infty}^{\infty} x(t)x(t-|\tau|)e^{-i2\pi f\tau}d\tau.\end{aligned}\tag{II.46}$$

Equation (II.46) offers a convenient method of computing the instantaneous power spectrum. Although rarely used, Turner's work produces insights into how the autocorrelation function relates to different time-frequency representations. Additionally, the non-uniqueness of Page's instantaneous power spectrum implies that many other instantaneous power spectra exist.

Levin's Instantaneous Power Spectrum

In 1964, Levin [Lev64] extended both Page's and Turner's works by developing a new instantaneous power spectrum and showing its equivalence to the ambiguity function which is a widely used technique in RADAR signal processing [Woo53]. Levin first defined two *running transforms* by

$$X_T^-(f) = \int_{-\infty}^T x(t)e^{-i2\pi ft}dt\tag{II.47}$$

and

$$X_T^+(f) = \int_T^{\infty} x(t)e^{-i2\pi ft}dt,\tag{II.48}$$

where $X_T^-(f)$ is the past data spectrum and $X_T^+(f)$ is the future data spectrum.

Levin then defined a past and a present instantaneous power spectrum as

$$\rho^-(T, f) = \frac{\partial}{\partial T} |X_T^-(f)|^2 \quad (\text{II.49})$$

and

$$\rho^+(T, f) = \frac{\partial}{\partial T} |X_T^+(f)|^2. \quad (\text{II.50})$$

Again, note that $\rho^-(T, f)$, (II.49), equals to Turner's instantaneous power spectrum. Levin then defined his instantaneous power spectrum by

$$\rho(T, f) = \frac{1}{2} \{ \rho^-(T, f) + \rho^+(T, f) \} \quad (\text{II.51})$$

which symmetrically treats past and future values of $x(t)$. Using a proof similar to Turner's, Levin showed

$$\begin{aligned} \rho(T, f) &= \text{Re} \left\{ x(T) e^{i2\pi f T} \int_{-\infty}^{\infty} x(t) e^{-i2\pi f t} dt \right\} \\ &= \text{Re} \{ x(T) e^{i2\pi f T} X(f) \}. \end{aligned} \quad (\text{II.52})$$

Equation (II.52) almost equals Page's instantaneous power spectrum for real signals, (II.29), except the Fourier transformation in (II.52) is computed over all time instead of over past values only.

Levin made two more interesting points. First, Levin related his instantaneous power spectrum to the ambiguity function. The ambiguity function represents a signal in the frequency-tau domain where tau is an autocorrelation lag number. The ambiguity function, used in tradeoff studies of random pulse range and velocity design [Bou83], is defined by

$$\chi(\theta, \tau) = \int_{-\infty}^{\infty} \hat{x}(t) \hat{x}(T + \tau) e^{i f T} dT, \quad (\text{II.53})$$

where $\hat{x}(t)$ is an analytic signal. Levin showed that

$$\int_{-\infty}^{\infty} dT \int_{-\infty}^{\infty} \rho(T, f) e^{i2\pi(\theta T + \tau f)} df = \chi^*(\theta, \tau). \quad (\text{II.54})$$

Thus, the two-dimensional Fourier transformation of Levin's instantaneous power spectrum equals the complex conjugate of the ambiguity function. The connection between the ambiguity function and Levin's instantaneous power spectrum is interesting because the ambiguity function equals the two-dimensional Fourier transformation of the Wigner distribution that will be described in this chapter.

Levin derived an alternate method for computing $\rho(t, f)$. First, he substituted the inverse Fourier transformation of $X(f)$ into (II.52) yielding

$$\begin{aligned} \rho(T, f) &= \text{Re} \left\{ x(T) e^{i2\pi f T} \int_{-\infty}^{\infty} x(\xi) e^{-i2\pi f \xi} d\xi \right\} \\ &= \text{Re} \left\{ \int_{-\infty}^{\infty} x(T) x(\xi) e^{i2\pi f (T - \xi)} d\xi \right\}. \end{aligned} \quad (\text{II.55})$$

Now if $\tau = \xi - T$, then

$$\rho(T, f) = \text{Re} \left\{ \int_{-\infty}^{\infty} x(T) x(T + \tau) e^{-i2\pi f \tau} d\tau \right\}. \quad (\text{II.56})$$

Comparing (II.56) with Turner's instantaneous power spectrum,

$$\rho(T, f) = \int_{-\infty}^{\infty} x(t) x(t - |\tau|) e^{-i2\pi f \tau} d\tau, \quad (\text{II.57})$$

indicates Levin's and Turner's instantaneous power spectra equal the Fourier transformation of an autocorrelation function signal. Chapter IV more intensely explores the relationship between the instantaneous power spectrum and the autocorrelation function.

Running Power Spectra and Autocorrelation Functions

All the time-frequency representations presented have, except for the spectrogram, involved the two-dimensional probability density function method. Since modifying the Fourier transformation yielded the spectrogram and limiting signals to present and to past values prompted the instantaneous power spectrum, Fano [Fan50], Schroeder and Atal [Sch62] combined these ideas to define a technique, like Page's, which uses only past signal values. However, they enhanced Page's work by noting the impracticality of using all past values since the analyzed signal may change globally.

Fano defined a running autocorrelation function as

$$\phi_t(\tau) = 2\alpha \int_{-\infty}^t x(\xi)x(\xi - \tau)e^{-2\alpha(t-\xi)}d\xi, \quad (\text{II.58})$$

where normally the time average autocorrelation function is computed by

$$\phi(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(\xi)x(\xi - \tau)d\xi \quad (\text{II.59})$$

and α is some constant scale factor. Thus, (II.58) and (II.59) differ in two respects. First, (II.58) does not use future signal values. Second, taking $t = 0$ and $\alpha = 1/2$, the $e^{-2\alpha(t-\xi)}$ term in (II.58) becomes e^{ξ} which is maximal for $\xi = 0$ and minimal for $\xi = -\infty$. Thus, the exponential term in (II.58) weights the current signal value the heaviest and the oldest signal term the lightest. Fano then defined the short-time power spectrum as

$$G_t(f) = 2\alpha \left| \int_{-\infty}^t x(\xi)e^{-i2\pi f\xi - \alpha(t-\xi)}d\xi \right|^2 \quad (\text{II.60})$$

and then related $G_t(f)$ to $\phi_t(\tau)$ by showing

$$\phi_i(\tau) = e^{\alpha\tau} \int_{-\infty}^{\infty} G_i(f) \cos(2\pi f\tau) df$$

$$\text{and} \quad G_i(f) = \int_{-\infty}^{\infty} \phi_i(\tau) e^{-\alpha\tau} \cos(2\pi f\tau) d\tau. \quad (\text{II.61})$$

The significance of this work is the relationship Fano established between the running autocorrelation function and the running spectrum. Again, a relationship between an autocorrelation and a spectral energy arises.

Schroeder and Atal, [Sch62], discussed and extended Fano's work by first noticing that (II.59) is not well-defined for negative τ 's since that would require future data knowledge. They defined the short-time autocorrelation function as

$$\phi(t, \tau) = \int_{-\infty}^t f(\xi) f(\xi - |\tau|) r_\phi(t - \xi) d\xi \quad (\text{II.62})$$

where $r_\phi(t)$ equals 0 for $t < 0$ and equals some weighting function for $t \geq 0$. The weighting function $r_\phi(t)$'s purpose is to weight current signal values more than previous signal values. Thus, the short-time autocorrelation function may be somewhat immune to slowly changing system characteristics. Schroeder and Atal also defined the short-time power spectrum as

$$G(f, t) = \left| \int_{-\infty}^t f(x) e^{-i2\pi f x} r_g(t - x) dx \right|^2. \quad (\text{II.63})$$

Unfortunately, no simple relationship between (II.62) and (II.63) holds unless

$$r_g(t) = e^{-\alpha t} \quad (\text{II.64})$$

$$r_\phi(t) = e^{-2\alpha t}. \quad (\text{II.65})$$

The short-time power spectrum (II.63) is very similar to the spectrogram except the spectrogram uses future signal values.

Rihaczek's Complex Energy Distribution

In 1968, Rihaczek related the works of Levin and Page with those of the Wigner distribution using the concept of complex energy [Rih68]. Complex energy values are counter-intuitive since energy normally is a real non-negative number as it is computed by the magnitude square of some complex number. However, nothing prevents defining a complex valued energy.

Rihaczek wanted energy estimates for a particular time and frequency. Using the fact that the instantaneous power can be written as

$$\varepsilon(t) = |x(t)|^2 = x(t)x^*(t) \quad (\text{II.66})$$

and using the Fourier transformation of $x(t)$, an energy estimate for time t is

$$\begin{aligned} \varepsilon(t) &= x(t) \left\{ \int_{-\infty}^{\infty} X(f) e^{i2\pi ft} df \right\}^* \\ &= x(t) \int_{-\infty}^{\infty} X^*(f) e^{-i2\pi ft} df. \end{aligned} \quad (\text{II.67})$$

Similarly, an instantaneous power can be written in terms of frequency by

$$\begin{aligned} \varepsilon(f) &= X(f)X^*(f) \\ &= \left\{ \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} dt \right\} X^*(f). \end{aligned} \quad (\text{II.68})$$

Looking at $\varepsilon(t)$ from (II.67), an estimate of $\varepsilon(t, f_0)$ for some Δf about f_0 can be defined as

$$\varepsilon(t, f_0) \approx x(t) \left\{ X(f_0) e^{-i2\pi f_0 t} \Delta f \right\}. \quad (\text{II.69})$$

Similarly from (II.68), an estimate of $\varepsilon(t_0, f)$ for some Δt about t_0 can be written

$$\varepsilon(t_0, f) \approx X^*(f) \left\{ x(t_0) e^{-i2\pi f t_0} \Delta t \right\}. \quad (\text{II.70})$$

Thus, one possible local energy estimate is

$$\varepsilon(t, f) \approx c x(t) X^*(f) e^{-i2\pi f t}. \quad (\text{II.71})$$

If $c = 1$, then

$$\varepsilon(t, f) \stackrel{\Delta}{=} x(t) X^*(f) e^{-i2\pi f t} \quad (\text{II.72})$$

defines the complex energy distribution. Note that $\varepsilon(t, f)$ can easily be complex valued since both $X^*(f)$ and $e^{-i2\pi f t}$ are complex valued.

Although the complex energy distribution $\varepsilon(t, f)$ was intuitively designed, does the complex energy distribution contain any interesting properties? The answer is yes, which can be seen from the following properties.

Property II.17: The marginal distribution with respect to t yields the Fourier transformation energy.

Proof:

$$\begin{aligned} \int_{-\infty}^{\infty} \varepsilon(t, f) dt &= \int_{-\infty}^{\infty} x(t) X^*(f) e^{-i2\pi f t} dt = X^*(f) \left\{ \int_{-\infty}^{\infty} x(t) e^{-i2\pi f t} dt \right\} \\ &= X^*(f) X(f) = |X(f)|^2. \end{aligned}$$

†††

Property II.18: The marginal distribution with respect to f yields the instantaneous power.

Proof:

$$\begin{aligned} \int_{-\infty}^{\infty} \varepsilon(t, f) df &= \int_{-\infty}^{\infty} x(t) X^*(f) e^{-i2\pi f t} df = x(t) \left\{ \int_{-\infty}^{\infty} X^*(f) e^{-i2\pi f t} df \right\} \\ &= x(t) \left\{ \int_{-\infty}^{\infty} X(f) e^{-i2\pi f t} df \right\}^* \\ &= x(t) x^*(t) = |x(t)|^2. \end{aligned}$$

†††

Property II.19: The Rihaczek distribution yields the total signal energy when integrated with respect to both variables.

Proof:

$$\begin{aligned}
 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varepsilon(t, f) dt df &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t) X^*(f) e^{-i2\pi f t} dt df \\
 &= \int_{-\infty}^{\infty} x(t) \left\{ \int_{-\infty}^{\infty} X^*(f) e^{-i2\pi f t} df \right\} dt \\
 &= \int_{-\infty}^{\infty} x(t) x^*(t) dt = \int_{-\infty}^{\infty} |x(t)|^2 dt = E_x.
 \end{aligned}$$

†††

Thus, the two marginal distributions yield the Fourier transformation energy and the instantaneous power, respectively, while the entire distribution integrates to the signal's total energy. Notice that Properties II.18 and II.19 compare exactly with the instantaneous power spectrum. Thus, the complex energy not only provides an intuitive energy measure, but it also matches the instantaneous power spectrum properties. The next section shows that the Wigner distribution also contains these three properties. Thus, the Rihaczek distribution generalizes Levin's, Page's and Wigner's techniques. The negative aspect of the Rihaczek distribution is the complex energy values which are difficult to interpret as spectral energy estimates.

The Rihaczek distribution is not widely used in practice because it contains complex values. Complex energy values are not easily interpreted when displayed on a console or in a graph because the modulus of complex valued functions must be computed. Thus, the Rihaczek's distribution cannot substitute for the spectrogram. Szu, [Szu84], claims that since the modulus function loses phase information, the nonstationary signal information is lost. However, Eggermont [Egg90] claims that the real part of the Rihaczek distribution contains time varying information because its imaginary part can be computed from its real part. Additionally, the main three Rihaczek properties can be found using only the real part of the Rihaczek distribution. Nonetheless, Rihaczek's distribution does produce some interesting concepts.

The Wigner Distribution

To this point, all the time-frequency representations discussed were based upon, at least in part, the Fourier transformation. A time-frequency representation based solely on a two-dimensional pdf approach is demonstrated by the Wigner distribution. In 1932 [Wig32], Wigner wanted a two-dimensional function, $W_x(\alpha, \beta)$, which operated on a one-dimensional signal, $x(\alpha)$, and would contain the following properties:

$$E = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_x(\alpha, \beta) d\alpha d\beta \quad (\text{II.73})$$

$$f_1(\alpha) = \int_{-\infty}^{\infty} W_x(\alpha, \beta) d\beta \quad (\text{II.74})$$

$$f_2(\beta) = \int_{-\infty}^{\infty} W_x(\alpha, \beta) d\alpha. \quad (\text{II.75})$$

Wigner's goal was to compute E , $f_1(\alpha)$, and $f_2(\beta)$ easily from $x(\alpha)$ by means of $W_x(\alpha, \beta)$.

The question was what should $W_x(\alpha, \beta)$ equal since an infinite number of functions, including Rihaczek's distribution, adhere to (II.73), (II.74), and (II.75). Arbitrarily, Wigner defined

$$W_x(\alpha, \beta) \triangleq \int_{-\infty}^{\infty} x(\alpha + \tau/2) x^*(\alpha - \tau/2) e^{-i2\pi\tau\beta} d\tau, \quad (\text{II.76})$$

which is now called the *Wigner distribution*. In the context of time-frequency representations, the Wigner distribution for a signal $x(t)$ is

$$W_x(t, f) = \int_{-\infty}^{\infty} x(t + \tau/2) x^*(t - \tau/2) e^{-i2\pi\tau f} d\tau. \quad (\text{II.77})$$

Chapter III will show that the Wigner distribution adheres to the three previous properties with the following equations

$$E_x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_x(t, f) dt df \quad (\text{II.78})$$

$$|x(t)|^2 = \int_{-\infty}^{\infty} W_x(t, f) df \quad (\text{II.79})$$

$$|X(f)|^2 = \int_{-\infty}^{\infty} W_x(t, f) dt, \quad (\text{II.80})$$

where E is the total energy, $|x(t)|^2$ is the instantaneous power, and $|X(f)|^2$ is the Fourier transformation energy. Thus, although the Rihaczek complex energy distribution and the Wigner distribution both adhere to these three properties, the Wigner distribution is not based upon extending the Fourier transformation. As such, the Wigner distribution is typically considered a two-dimensional pdf (within a scale factor of E) that indicates where signal energy occurs in time and frequency.

Although the Wigner distribution does not directly extend the Fourier transformation, the Wigner distribution provides some kind of spectral energy estimates. For cases when the Wigner distribution produces poor spectral energy estimates, the marginal distributions may provide time varying signal energy information. For these reasons, the Wigner distribution is becoming widely used in situations where signals vary with time. In fact, the Wigner distribution is probably the second most utilized time varying spectral analysis tool, behind the spectrogram. However, the Wigner distribution has limitations which will be discussed in Chapter III.

The Wavelet Transformation

All the time-frequency representations discussed so far follow either the Fourier transformation extension method or the two-dimensional pdf method. The third time varying spectral energy technique, alternate basis functions, has not yet been discussed.

Two reasons exist for waiting until now to discuss a technique which uses an alternate basis function. First, the other two methods produce usable techniques; the techniques may not be perfect, but the techniques may provide valuable time varying spectral information. Second, although alternate basis function techniques produce more powerful analysis tools, the techniques typically require much more computing time. Additionally, if non-orthogonal alternate basis functions arise, interpreting the time-frequency domain becomes more difficult since Parseval's theorem no longer holds. In spite of these difficulties, currently there exists a surge in researching new basis functions for time varying spectral analysis. This research investigates techniques that produce time varying spectral energies capable of replacing the spectrogram. The most popular alternate basis, the wavelet transformation, is described next.

The wavelet transformation dispels the notion of time-frequency representations with a time-scale representation. As such, wavelet transformations are different from the other time varying signal analysis tools. However, wavelet transformations parallel the short-time Fourier transformation in some regards. The wavelet transformation and the short-time Fourier transformation connection is seen by rewriting the short-time Fourier transformation as

$$X(t, f) = \int_{-\infty}^{\infty} x(\xi) h(\xi - t) e^{-i2\pi f \xi} d\xi, \quad (\text{II.81})$$

where the only difference between (II.8) and (II.81) is the position of $x(\xi)$ and $h(\xi - t)$.

Next, let

$$\Phi_{t,f}(\xi) = h(\xi - t) e^{-i2\pi f \xi} \quad (\text{II.82})$$

so that

$$X(t, f) = \int_{-\infty}^{\infty} x(\xi) \Phi_{t,f}(\xi) d\xi. \quad (\text{II.83})$$

Equation (II.83) is just a different way to look at the short-time Fourier transformation. However, (II.83) can also be taken as a inner-product between $x(\xi)$ and $\Phi_{t,f}(\xi)$. As such, (II.83) describes the projection of $x(\xi)$ into the subspace defined by $\Phi_{t,f}(\xi)$. Thus, (II.83) decomposes $x(\xi)$ in terms of $\Phi_{t,f}(\xi)$. The term $\Phi_{t,f}(\xi)$ implies that the window function acts upon the complex exponential in the same fashion as it acts upon signals. That is, windowing the complex exponential limits the time duration of the decomposing signals. Therefore, (II.83) decomposes $x(\xi)$ into time limited sinusoids so that the basis function now depends on time and frequency. Basis functions depending on both time and frequency differ from the complex exponential basis functions that depend only on frequency. Thus, the spectrogram can be considered a time-frequency representation based upon the alternate basis function method.

Similar to the spectrogram, a *wavelet transformation* is defined as

$$C_{a,b} = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \Phi_{a,b}(t) dt, \quad (\text{II.84})$$

where $\Phi_{a,b}(t)$ is a basis function. The new basis function, $\Phi_{a,b}(t)$, is called a *wavelet*. A wavelet is a time dependent oscillating function whose frequency is inversely proportional to the dilation parameter a and whose time is the shift parameter b . The non-parameterized wavelet, $\Phi(t)$, is a complete basis function if $\Phi(t)$ is absolutely integrable, square integrable, and has a zero mean. Thus, a parameterized wavelet function, $\Phi_{a,b}(t)$, must be absolutely integrable, square integrable, and have zero mean for all a and b . An example of a general wavelet function is

$$\Phi(t) = e^{-\frac{t^2}{2}} e^{i2\pi ft}. \quad (\text{II.85})$$

A two parameter example, $\Phi_{a,b}(t)$, can be computed using

$$\begin{aligned}\Phi_{a,b}(t) &= \Phi\left[\frac{t-b}{a}\right] \\ &= e^{-\frac{(t-b)^2}{a^2}} e^{i2\pi\frac{f}{a}(t-b)},\end{aligned}\tag{II.86}$$

where f/a is the wavelet frequency.

Rewriting (II.82) with the Gaussian window function yields

$$\Phi_{t,f}(\xi) = e^{-\frac{\sigma^2}{2}} e^{-i2\pi f\xi}.\tag{II.87}$$

Letting $t = 0$ and $a = 1$ for the wavelet and letting $t = 0$ and $\sigma = 2$ for the short-time Fourier transformation produces

$$\Phi_{1,b}(0) = e^{-b^2} e^{-i2\pi fb}$$

and

$$\Phi_{0,f}(\xi) = e^{-\xi^2} e^{-i2\pi f\xi}.\tag{II.88}$$

Thus, the wavelet transformation is identical to the short-time Fourier transformation for the given constants. However, notice one additional variable, $a = 1$, exists in the wavelet transformation and not in the short-time Fourier transformation. The variable a is the dilation parameter that controls the length of the decomposing basis function while simultaneously controlling the basis function frequency. The additional degree of freedom of the wavelet transformation produces interesting new spectral energy information that is being investigated by many researchers.

Cohen's General Time-Frequency Representation

The previous sections described the more popular time-frequency representations proposed over the years. Comparing these different time-frequency representations is difficult because each technique typically works well only for specific signal types.

Thus, techniques that analyze properties of all time-frequency representations within the same context are desirable. With such a technique, different properties could compare competing time-frequency representations. Fortunately, such a technique exists.

In 1966, Cohen [Coh66] defined the following general time-frequency representation

$$P_x(t, f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(\theta, \tau) x\left(u + \frac{\tau}{2}\right) x^*\left(u - \frac{\tau}{2}\right) e^{-i2\pi(\theta t + \tau f - \theta u)} du d\tau d\theta \quad (\text{II.89})$$

which describes many of the popular time-frequency representations by assigning the kernel function, $\phi(\theta, \tau)$, different functions. For instance the following properties hold:

Property II.20: If $\phi(\theta, \tau) = 1$, then $P_x(t, f)$ is the Wigner distribution.

Property II.21: If $\phi(\theta, \tau) = e^{i\theta\tau/2}$, then $P_x(t, f)$ is the Rihaczek distribution.

Property II.22: If $\phi(\theta, \tau) = e^{i\theta|\tau|/2}$, then $P_x(t, f)$ is the Page distribution.

Property II.23: If $\phi(\theta, \tau) = \int_{-\infty}^{\infty} h(\xi + \tau/2) h^*(\xi - \tau/2) e^{-i\theta\xi} d\xi$, then $P_x(t, f)$ is the spectrogram.

Thus, (II.89) generalizes time-frequency representations into one form. This generalization allows analyzing different time-frequency distribution characteristics solely on the kernel function, $\phi(\theta, \tau)$. As an example, Cohen showed that three time-frequency representation characteristics

$$|x(t)|^2 = \int_{-\infty}^{\infty} P_x(t, f) df \quad (\text{II.90})$$

$$|X(f)|^2 = \int_{-\infty}^{\infty} P_x(t, f) dt \quad (\text{II.91})$$

$$E = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_x(t, f) dt df \quad (\text{II.92})$$

can be seen by using the following general time-frequency properties:

$$\text{Property II.24:} \quad \phi(\theta, 0) = 1 \Leftrightarrow |x(t)|^2 = \int_{-\infty}^{\infty} P_x(t, f) df.$$

$$\text{Property II.25:} \quad \phi(0, \tau) = 1 \Leftrightarrow |X(f)|^2 = \int_{-\infty}^{\infty} P_x(t, f) dt.$$

$$\text{Property II.26} \quad \phi(0, 0) = 1 \Leftrightarrow \text{total energy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_x(t, f) dt df.$$

For example, the three time-frequency representation properties hold for the Wigner distribution since

$$\phi(\theta, \tau) = 1 \Rightarrow \phi(\theta, 0) = \phi(0, \theta) = \phi(0, 0) = 1. \quad (\text{II.93})$$

For the Rihaczek distribution,

$$\begin{aligned} \phi(\theta, \tau) &= e^{i\theta\tau/2} \Rightarrow \phi(\theta, 0) = e^0 = 1 \\ \phi(0, \tau) &= e^0 = 1 \\ \phi(0, 0) &= e^0 = 1 \end{aligned} \quad (\text{II.94})$$

validates the three properties. Similarly, Page's instantaneous power spectrum contains these properties. The spectrogram however maintains the three properties only under strict conditions. Using the spectrogram kernel, notice that

$$\phi(\theta, 0) = \int_{-\infty}^{\infty} h(\xi) h^*(\xi) e^{-i\theta\xi} d\xi = 1 \quad (\text{II.95})$$

$$\text{only if} \quad h(\xi) h^*(\xi) = \delta(\xi), \quad (\text{II.96})$$

where $\delta(\xi)$ is the Dirac delta function. Because a delta function window is impractical, the spectrogram does not maintain the first property. Thus, the spectrogram marginal distribution with respect to frequency does not equal the instantaneous power. Again, the general time-frequency properties produce

$$\phi(0, \tau) = \int_{-\infty}^{\infty} h(\xi + \tau/2) h^*(\xi - \tau/2) e^{-i\theta(0)} d\xi = 1 \quad (\text{II.97})$$

only if
$$h(\xi + \tau/2) h^*(\xi - \tau/2) = \delta(\xi). \quad (\text{II.98})$$

Since most real world windows are symmetric, we can assume $h(\xi) = h(-\xi)$ so that

$$h(\tau/2) = h(-\tau/2) = 1. \quad (\text{II.99})$$

The rectangular window, $h(\tau/2) = 1$, allows the spectrogram marginal distribution with respect to time to equal the Fourier transformation energy. Thus, an impulse window allows computing the instantaneous power while a rectangular window allows obtaining the Fourier transformation energy. For the total energy property,

$$\phi(0, 0) = \int_{-\infty}^{\infty} h(\xi) h^*(\xi) d\xi = \int_{-\infty}^{\infty} |h(\xi)|^2 d\xi = 1 \quad (\text{II.100})$$

implies that if the squared window area equals 1, then the spectrogram contains a signal's total energy.

Another interesting property that Cohen's generalized time-frequency representation allows verifying whether a time-frequency representation is real valued or not.

Cohen showed [Coh89]

$$\phi(\theta, \tau) = \phi^*(-\theta, -\tau) \Leftrightarrow P_x(t, f) \in \mathfrak{R}. \quad (\text{II.101})$$

For the Wigner distribution,

$$\phi(\theta, \tau) = 1 = 1 = \phi(-\theta, -\tau) \quad (\text{II.102})$$

implies real Wigner distribution values. Similarly, Page's kernel yields

$$\phi(\theta, \tau) = e^{i\theta\tau/2} = e^{-i(-\theta)(-\tau)/2} = \phi^*(-\theta, -\tau). \quad (\text{II.103})$$

Thus, Page's instantaneous power spectrum values are real. The Rihaczek distribution does not behave so nicely since

$$\phi(\theta, \tau) = e^{i\theta\tau/2} \neq e^{-i(-\theta)(-\tau)/2} = \phi^*(-\theta, -\tau). \quad (\text{II.104})$$

Thus, the Rihaczek distribution values are not real. Finally, the spectrogram kernel produces

$$\phi(\theta, \tau) = \int_{-\infty}^{\infty} h(\xi + \tau/2) h^*(\xi - \tau/2) e^{-i\theta\xi} d\xi$$

and

$$\begin{aligned} \phi^*(-\theta, -\tau) &= \left(\int_{-\infty}^{\infty} h(\xi - \tau/2) h^*(\xi + \tau/2) e^{-i(-\theta)\xi} d\xi \right)^* \\ &= \int_{-\infty}^{\infty} h^*(\xi - \tau/2) h(\xi + \tau/2) e^{-i\theta\xi} d\xi, \end{aligned} \quad (\text{II.105})$$

which indicates that the spectrogram is real for all windows.

Cohen's general time-frequency representation allows investigating time-frequency properties for a wide range of techniques. Several of these properties were proved for the main time-frequency representations. Additionally, Cohen's general time-frequency representation indicates that filtering the Wigner distribution generates all other time-frequency representations because the kernel function, $\phi(\theta, \tau)$, can be considered filtering the integrand of (II.89). Unfortunately, Cohen's general time frequency representation does not indicate whether a particular time-frequency representation better estimates time varying spectral energies when compared to the spectrogram. In Chapter IV, a new general approach, called the correlation approach to time-frequency representations is presented, which helps determine if a particular time-frequency representation can replace the spectrogram. This is one of the main results of this research.

Chapter Summary

This chapter presented many different techniques that describe time varying signals with respect to time and frequency. The goal of these techniques is to estimate spectral energy as a function of time. Ideally, these techniques should allow replacing the spectrogram with the time-frequency representation. Unfortunately, these spectral energy

estimates may contain anomaly signals, complex values, or negative values. Thus, signal processing may have to use these time-frequency representations indirectly instead of directly. Many times the marginal distributions may provide information as to how signal energy changes over time.

The goal of this thesis is to develop new techniques that analyze whether alternate time varying spectral energy techniques can replace the spectrogram directly or indirectly. While working on this goal, two discoveries were made. First, a new method for analyzing time-frequency representations was developed. This method is thoroughly discussed in Chapter IV. The second discovery, discussed in Chapter V, results from attempting to compute spectral energy information from any time-frequency representation.

Before discussing the new discoveries, a complete overview of the Wigner distribution is presented in Chapter III. Why discuss the Wigner distribution thoroughly and not the other time-frequency representations? Well, since Cohen's general time-frequency representation indicates that the Wigner distribution generates other time-frequency representations, better understanding the Wigner distribution produces further understanding of all time-frequency representations. Additionally, many researchers present the Wigner distribution as a spectrogram replacement ignoring or dismissing the Wigner distribution anomalies as insignificant. This viewpoint is incorrect and has unnecessarily stagnated research into other time-frequency representations. Chapter III shows the significance of Wigner distribution anomalies and also shows why the Wigner distribution may be important to time varying spectral energy analysis.

CHAPTER III

THE WIGNER DISTRIBUTION

Introduction

Chapter II reviewed several techniques that estimate spectral energy as a function of time. Although many of these techniques contain interesting properties, signal processing algorithms typically only use the spectrogram and the Wigner distribution. The spectrogram handles time varying signals by directly extending Fourier transformation energy. However, the spectrogram performs poorly unless the signal is short-time time-invariant; even then, the spectrogram produces poor spectral energy estimates during transition regions. Because transition regions contain strongly time varying signals, the spectrogram's window size would be so short that no usable spectral information could be obtained. Thus, robust time varying spectral energy techniques require using alternate time varying spectral energy techniques. The Wigner distribution, one of the many time-frequency representations, is under study by a number of researchers as an alternative to spectrograms. Because the Wigner distribution is so popular, this chapter overviews Wigner distribution properties and shows how these properties effect implementing the Wigner distribution.

When applying time varying spectral analysis techniques to signal processing, Chapter II stated that many researchers replace the spectrogram with the Wigner distribution and dismiss the Wigner distribution anomalies as insignificant. Thus, the Wigner distribution is often misused. This chapter presents examples of when the Wigner distribution performs better than the spectrogram and gives examples of when the Wigner

distribution performs worse than the spectrogram. However, even when the Wigner distribution cannot replace the spectrogram for estimating time varying spectral energies, techniques may exist for extracting spectral energy information from the Wigner distribution. This chapter ends by providing examples of extracting spectral energy information on real signals using the Wigner distribution.

Wigner Distribution Theory

Origins of the Wigner Distribution

In 1932, Eugene Wigner [Wig32] developed a method which replaced the operator theory method for calculating the position and momentum in quantum mechanical systems. Wigner's technique computed average positions and momentums using expected values. Specifically, Wigner defined

$$F_w(q, p, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi^*(q - 1/2\tau \hbar, t) e^{-i\tau p} \psi(q + 1/2\tau \hbar, t) d\tau \quad (\text{III.1})$$

where q is position, p is momentum, \hbar is Plank's constant, t is time and ψ is a wave function. Integrating (III.1) with respect to q or p yields marginal distributions that produce correct quantum mechanical distributions. Although this formulation makes little sense in the context of signal processing, the idea of generating a two-dimensional distribution function specifically to contain useful marginal distributions was unique [Coh66].

In 1948, while investigating time varying signals in communication theory, Ville [Vil48] applied the Wigner distribution to signal analysis. Ville replaced the position and momentum variables of (III.1) with time and frequency variables. This substitution produced

$$W_x(t, f) = \int_{-\infty}^{\infty} x(t + \tau/2) x^*(t - \tau/2) e^{-i2\pi f \tau} d\tau, \quad (\text{III.2})$$

where $x(t)$ represents any real or complex valued signal and $W_x(t, f)$ is the Wigner distribution. Ville applied the Wigner distribution to signal processing because the Wigner distribution contained certain desirable properties. These are described in the next section. The properties allow for successful implementations of the Wigner distribution to communication theory. However, very little research work was done on the Wigner distribution and its applications to signal processing until Classen and Mecklenbrauker formulated a discrete Wigner distribution in 1980, [Cla80a] and [Cla80b].

Wigner Distribution Properties

As discussed in Chapter II, the Wigner distribution represents a one-dimensional signal in two dimensions such that

$$\text{Property III.1:} \quad |x(t)|^2 = \int_{-\infty}^{\infty} W_x(t, f) df$$

$$\text{Property III.2:} \quad |X(f)|^2 = \int_{-\infty}^{\infty} W_x(t, f) dt$$

$$\text{Property III.3:} \quad E_x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_x(t, f) dt df.$$

Since Properties III.1 through III.3 form the basis of the Wigner distribution, the properties' validity is shown. Substituting the Wigner distribution equation, (III.2), into

Property III.1 produces

$$\int_{-\infty}^{\infty} W_x(t, f) df = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} x(t + \tau/2) x^*(t - \tau/2) e^{-i2\pi f\tau} d\tau \right\} df. \quad (\text{III.3})$$

Rearranging the integral of (III.3) yields

$$\begin{aligned}
\int_{-\infty}^{\infty} W_x(t, f) df &= \int_{-\infty}^{\infty} x(t - \tau/2) x^*(t - \tau/2) \left\{ \int_{-\infty}^{\infty} e^{-i2\pi f \tau} df \right\} d\tau \\
&= \int_{-\infty}^{\infty} x(t + \tau/2) x^*(t - \tau/2) \delta(\tau) d\tau \\
&= x(t) x^*(t) = |x(t)|^2.
\end{aligned} \tag{III.4}$$

Thus, the Wigner distribution marginal distribution, with respect to frequency, produces the instantaneous power. Similarly, substituting (III.1) into Property III.2 yields

$$\int_{-\infty}^{\infty} W_x(t, f) dt = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} x(t + \tau/2) x^*(t - \tau/2) e^{-i2\pi f \tau} d\tau \right\} dt. \tag{III.5}$$

Setting

$$t_1 = t - \tau/2 \text{ and } t_2 = t + \tau/2 \tag{III.6}$$

implies that

$$\tau = t_2 - t_1 \text{ and } t = (t_1 + t_2)/2 \tag{III.7}$$

and

$$\left| \frac{\partial(\tau, t)}{\partial(t_1, t_2)} \right| = 1 \tag{III.8}$$

so that

$$\begin{aligned}
\int_{-\infty}^{\infty} W_x(t, f) dt &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t_2) x^*(t_1) e^{i2\pi f(t_2 - t_1)} dt_1 dt_2 \\
&= \left\{ \int_{-\infty}^{\infty} x(t_2) e^{-i2\pi f t_2} dt_2 \right\} \left\{ \int_{-\infty}^{\infty} x^*(t_1) e^{i2\pi f t_1} dt_1 \right\} \\
&= X(f) \left\{ \int_{-\infty}^{\infty} x(t_1) e^{-i2\pi f t_1} dt_1 \right\}^* \\
&= X(f) X^*(f) = |X(f)|^2.
\end{aligned} \tag{III.9}$$

Thus, the Wigner distribution marginal distribution with respect to time equals the Fourier transformation energy. Substituting Property III.1 into Property III.3 proves Property III.3 by yielding

$$\begin{aligned}
E_x &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} W_x(t, f) df \right\} dt \\
&= \int_{-\infty}^{\infty} |x(t)|^2 dt,
\end{aligned} \tag{III.10}$$

which equals the total energy. Thus, integrating the Wigner distribution with respect to both variables, yields the signal's total energy.

The remaining Wigner distribution properties did not originate from the initial design constraints but are by-products of the two-dimensional representation. Thus, these properties are listed without proof since their derivations are readily found in [Bou83] and since applying the properties is more important than proving the properties. The following consist of the more important Wigner distribution properties:

Property III.4: $W_x(t, f) \in \Re$

Property III.5: $x(t) = 0, \forall |t| > T \Rightarrow W_x(t, f) = 0, \forall |t| > T$

Property III.6: $X(f) = 0, \forall |f| > \Omega \Rightarrow W_x(t, f) = 0, \forall |f| > \Omega$

Property III.7: $y(t) = x(t - t_0) \Rightarrow W_y(t, f) = W_x(t - t_0, f)$

Property III.8: $Y(f) = X(f - f_0) \Rightarrow W_y(t, f) = W_x(t, f - f_0)$

Property III.9: $x(t) \otimes y(t) \Rightarrow W_x(t, f) \otimes_t W_y(t, f) = \int_{-\infty}^{\infty} W_x(t, f) \cdot W_y(t - \xi, f) d\xi$

Property III.10: $x(t) \cdot y(t) \Rightarrow W_x(t, f) \otimes_f W_y(t, f) = \int_{-\infty}^{\infty} W_x(t, f) \cdot W_x(t, f - \xi) d\xi$

Property III.11: $x(t) = \frac{1}{x^*(0)} \int_{-\infty}^{\infty} W_x(t/2, f) e^{i2\pi ft} df.$

Since these eleven properties allow applying the Wigner distribution to signal processing, each property is discussed and related to the ideal properties discussed in Chapter II. Property III.4 states that the Wigner distribution always produces real values. Recall that two of the time-frequency representation methods, the Fourier transformation extension method and the two-dimensional pdf method, expect real values. The reason that real values are so important is that spectral energy (the square of a modulus function) is always real valued. Thus, any spectral energy representation that outputs complex values, like the Rihaczek distribution, disallows directly interpreting time-frequency coefficients as spectral energy values. Therefore, because the Wigner distribution produces real values, the Wigner distribution may be interpreted as a true spectral energy function.

The finite time support and finite frequency support properties of the Wigner distribution result from Properties III.5 and III.6, respectively. The importance of these two properties can be seen from Figure III.1 which indicates how these properties relate to time-frequency representations. Figure III.1 shows two graphs on the left side that indicate where $x(t)$ and $X(f)$ are non-zero. Figure III.1 further shows a graph on the right side that indicates non-zero time-frequency domain values. The reader should note that finite time support and finite frequency support simultaneously can never occur. However, Figure III.1 shows the ramifications of the case when either a signal contains finite time support or finite frequency support. Any signal processing algorithm that searches for the presence of a signal should either look at time based energy or frequency based energy signal representations. Thus, a time-frequency representation that adheres to Property III.5 allows finding the non-zero signal values directly. Similarly, a time-frequency representation that adheres to Property III.6 allows identifying a signal characterized by a maximum frequency, say f_1 , by searching the time-frequency domain

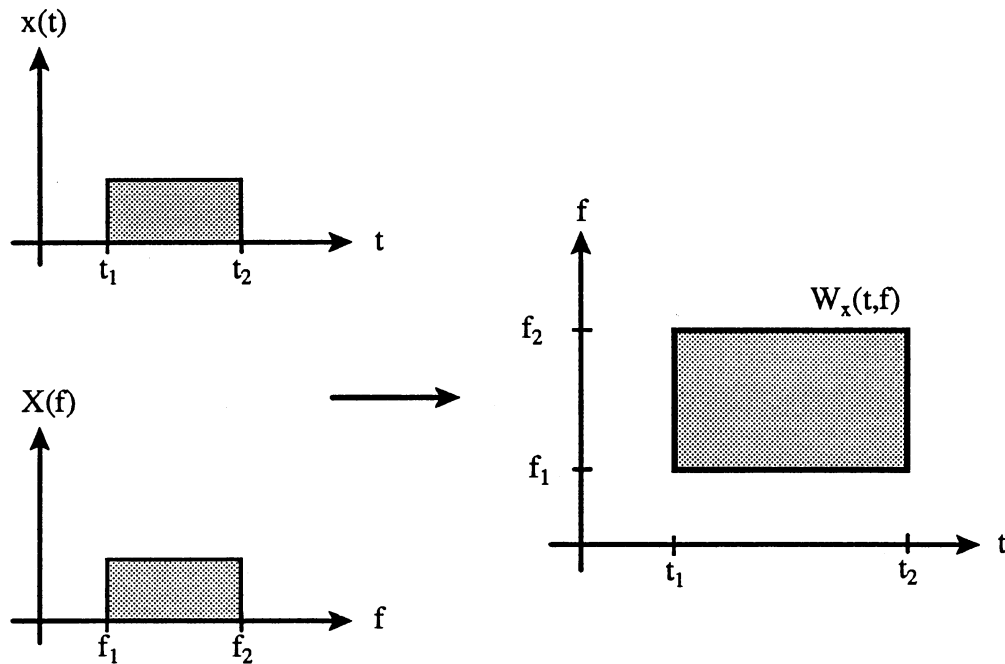


Figure III.1 An Example of Time and Frequency Support

for the presence of f_1 . Additionally, Property III.6 is helpful if plotting the Wigner distribution for a particular time is desired since graphs without this property would be difficult to analyze.

Properties III.7 and III.8 show that the Wigner distribution is shift-invariant in both time and frequency. The shift-invariant property is critical for identifying when a signal occurs and what frequencies comprise a signal. Additionally, recall that the relationships between convolving in the time domain and multiplying in the frequency domain relied on a linear time-invariant system. The reliance on a linear time-invariant system in terms

of the Wigner distribution partially depends on the shift-invariant property.

Property III.9 relates convolving in the time domain to the Wigner distribution. Recall that for linear time-invariant systems, convolving two signals $x(t)$ and $y(t)$ results in multiplying their Fourier transformations, $X(f)$ and $Y(f)$. The time convolution and frequency multiplication does not hold for the Wigner distribution, however. Property III.9 indicates that the Wigner distribution maintains a similar relationship to linear time-invariant systems with respect to frequency, but Property III.9 includes an additional convolution of Wigner distributions with respect to the time variables. Although the additional Wigner distribution convolution with respect to time is not entirely obvious, convolved time domain signals should transfer to the resulting Wigner distribution. Thus, Property III.9 indicates that convolving two time domain signals produces the expected effects in the corresponding Wigner distributions. Similarly, Property III.10 indicates that multiplying two time domain signals results in convolving their respective Wigner distributions with respect to frequency. Therefore, the basic relationship of linear time-invariant systems involving convolution and multiplication is maintained by the Wigner distribution.

The last property listed, Property III.11, indicates that completely retrieving a signal $x(t)$ from its Wigner distribution is possible up to a multiplicative constant. Thus, although a signal cannot be completely resolved from its Wigner distribution, storing the signal's value at time 0 along with the signal's Wigner distribution allows recovering the exact signal. Correctly obtaining a signal from its Wigner distribution implies that signal processing may use the Wigner distribution for coding applications. Wigner distribution based coding algorithms might reduce storage requirements for time varying signal since the Wigner distribution contains time varying spectral energy information.

All these properties indicate that the Wigner distribution contains some useful qualities. In fact, several of these properties should allow treating the Wigner distribution like an ordinary spectral energy. The best way to determine whether the Wigner distribution

behaves like a spectral energy is to compare Wigner distributions with spectrograms over a variety of signals. Thus, several examples are presented to indicate how well the Wigner distribution performs as a spectral estimator when compared to the spectrogram. Note that all spectra are computed using Gaussian windows.

The first example is a single sinusoid defined as

$$x(t) = \sin(2\pi ft), \quad (\text{III.11})$$

for f equalling 33.9 Hz. Although frequency f in (III.11) does not vary with time, the Wigner distribution should produce reasonable spectral estimates; otherwise, the Wigner distribution would not be very useful. The Wigner distribution and spectrogram contours, along with the signal graph, are given in Figure III.2. Recall from Chapter I that the contours indicate the presence of energy by plotting a black dot at the appropriate time and frequency location for a particular energy level. Thus, both contours indicate spectral energy for all time and at $f = 0.13$. Both spectral energy techniques compare favorably since both techniques indicate the presence of the sinusoid. Figure III.3 shows the spectrogram and Wigner distribution spectral energy estimates for $t = .5$ sec. The energy spectral estimates reinforce the claim that the two techniques produce similar results. However, notice that the Wigner distribution produces a wider and a noisier spectral energy estimate when directly compared with the spectrogram based spectral energy estimate. In addition, note that both spectral energy estimates should be impulse functions, but spectral leakage from the Gaussian windows prevents this from happening. Also, notice that the sinusoid in Figure III.2 contains non-uniform peaks. These nonuniform peaks arise from the sampling rate of the signal and do add some spectral leakage. However, the spectral leakage of the Gaussian window overwhelms the nonuniform peak spectral leakage.

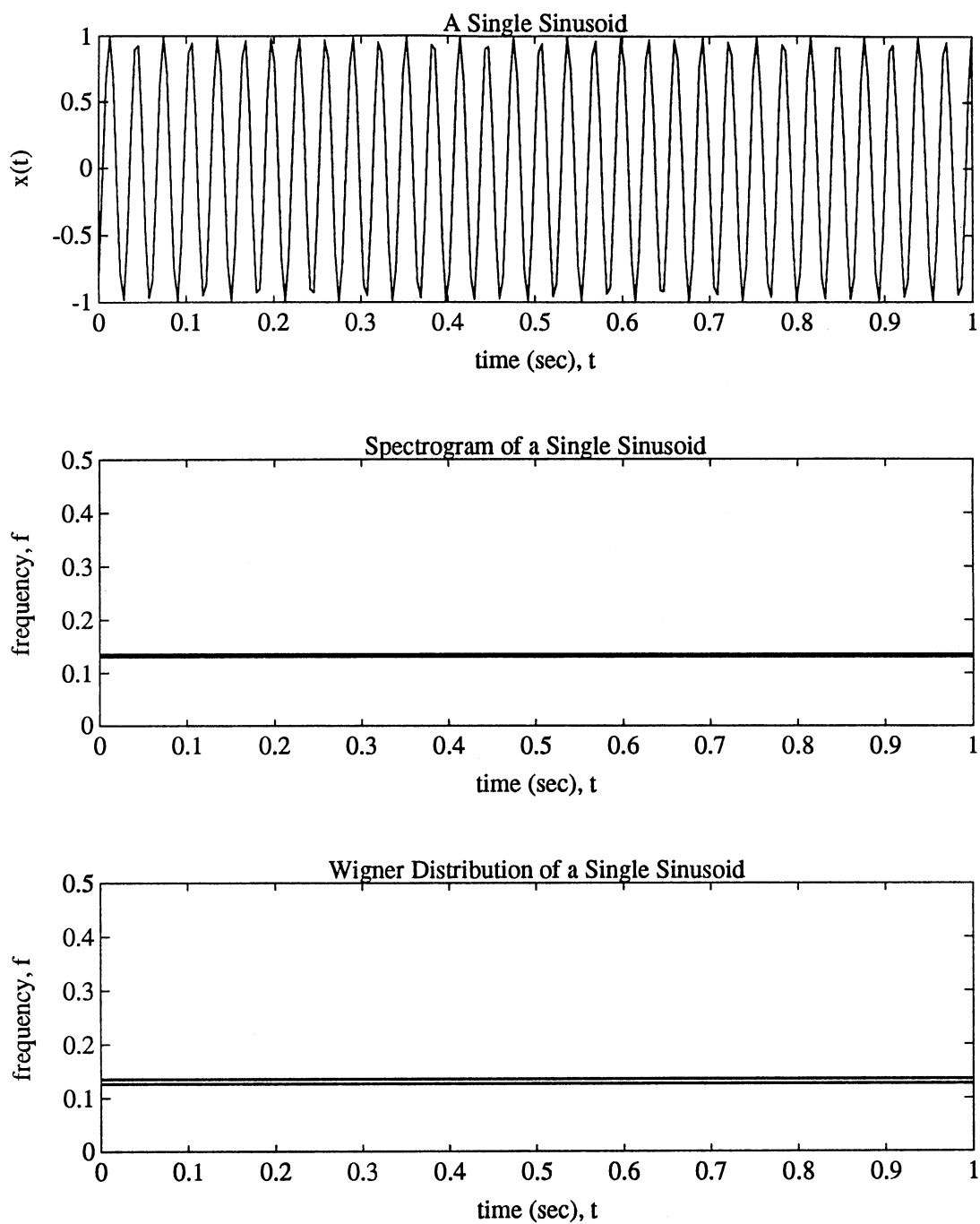


Figure III.2 The Spectrogram and Wigner Distribution of a Simple Sinusoid

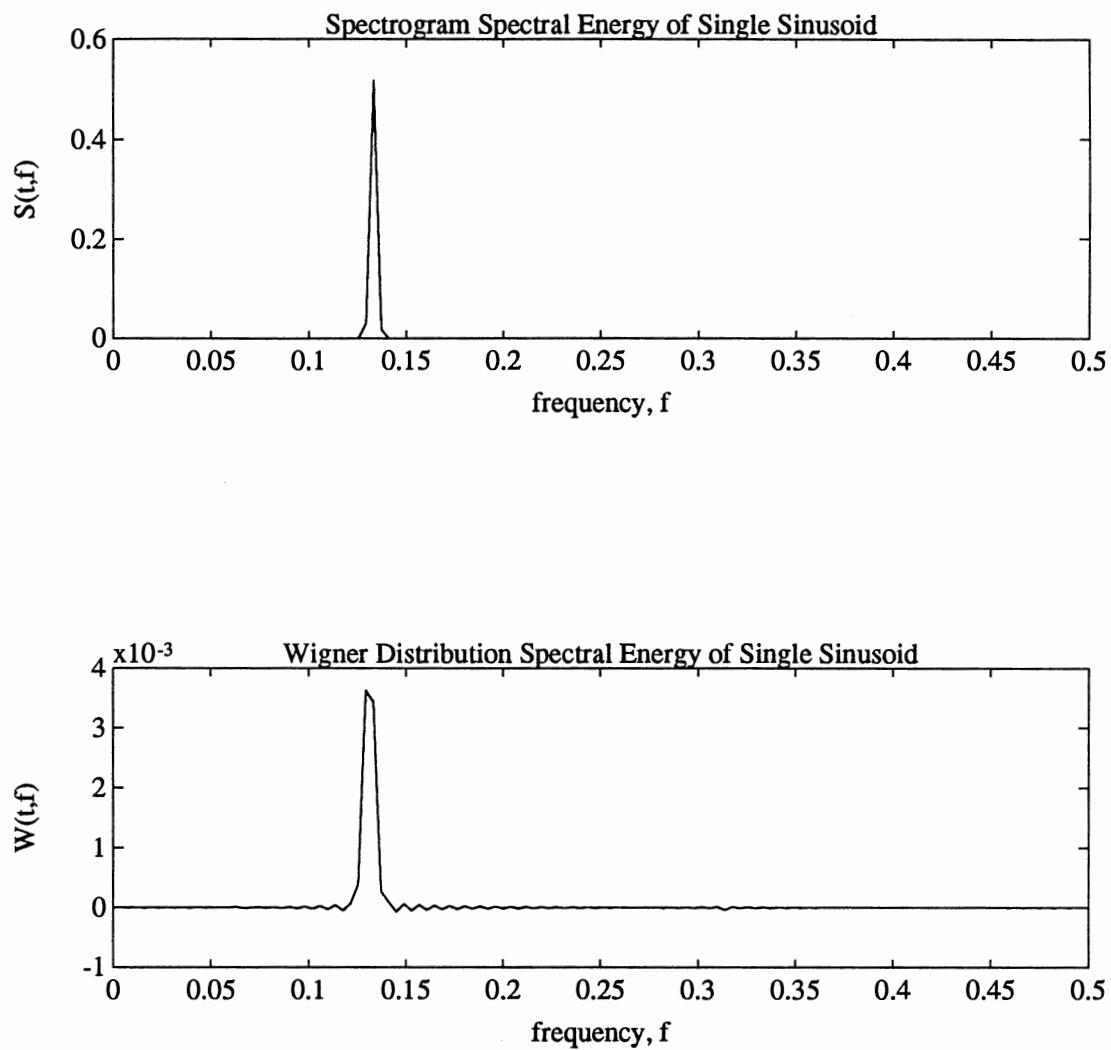


Figure III.3 Spectrogram and Wigner Distribution Spectral Energy Estimates of a Simple Sinusoid ($t = 0.5$ secs)

The next example probes the Wigner distribution performance on the simplest time varying signal, a chirp signal. The chirp signal is expressed by

$$x(t) = \sin(2\pi f(t)t), \quad (\text{III.12})$$

where $f(t)$ varies linearly from 33.9 Hz to 157 Hz. Figure III.4 presents the chirp signal, the chirp signal spectrogram, and the chirp signal Wigner distribution. The contours indicate that both the spectrogram and the Wigner distribution estimate the increasing frequency over time. However, the Wigner distribution produces more accurate estimates. Since the spectral energy estimate ideally is an impulse function, Figure III.5, which displays the spectrogram and Wigner distribution spectral energy estimates for $t = 0.5$ sec, shows that the Wigner distribution outperforms the spectrogram. Thus, the Wigner distribution actually produces better spectral energy estimates for chirp signals when compared to the spectrogram. This example has spawned many research claims of the Wigner distribution's superiority over the spectrogram for time varying signals. However, performance on one specific signal does not validate any general claims. Thus, additional examples are presented.

The third example extends the first example by adding an additional sinusoid to (III.11). The new signal is defined by

$$x(t) = \sin(2\pi f_1 t) + \sin(2\pi f_2 t), \quad (\text{III.13})$$

where $f_1 = 33.9$ Hz and $f_2 = 77.3$ Hz. Figure III.6 shows the two sinusoid signal, the spectrogram, and the Wigner distribution. Notice that the spectrogram and the Wigner distribution compare favorably at f_1 and f_2 (the different Wigner distribution estimates for f_1 and f_2 is purely due to plotting resolution, they actually are identical), but the Wigner distribution produces artifacts at 2.1 radians/second or 49.5 Hz. The artifact, called a *crossterm*, is directly related to the $x(t + \tau/2) \cdot x^*(t - \tau/2)$ term of the Wigner distribution.

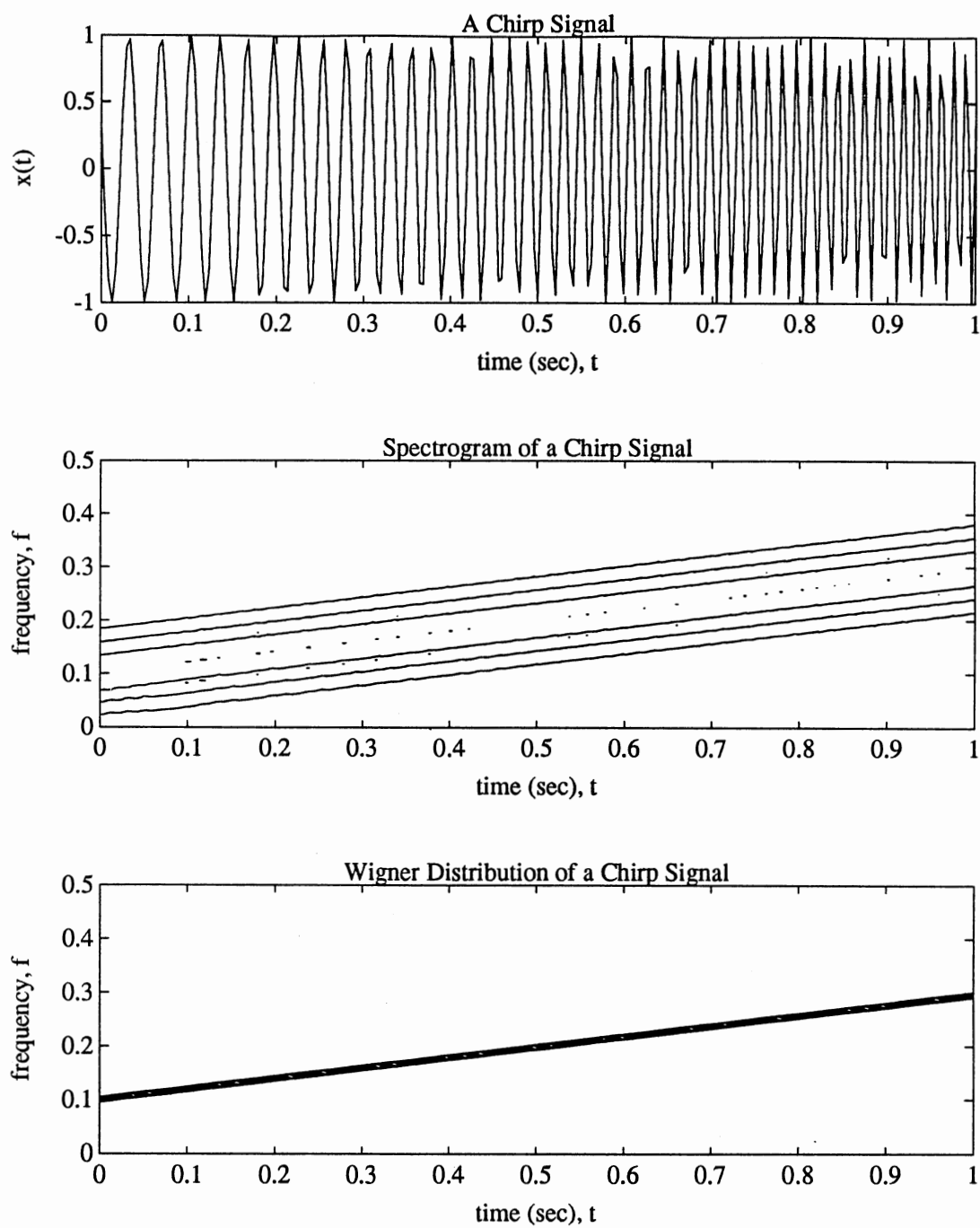


Figure III.4 The Spectrogram and Wigner Distribution of a Chirp Signal

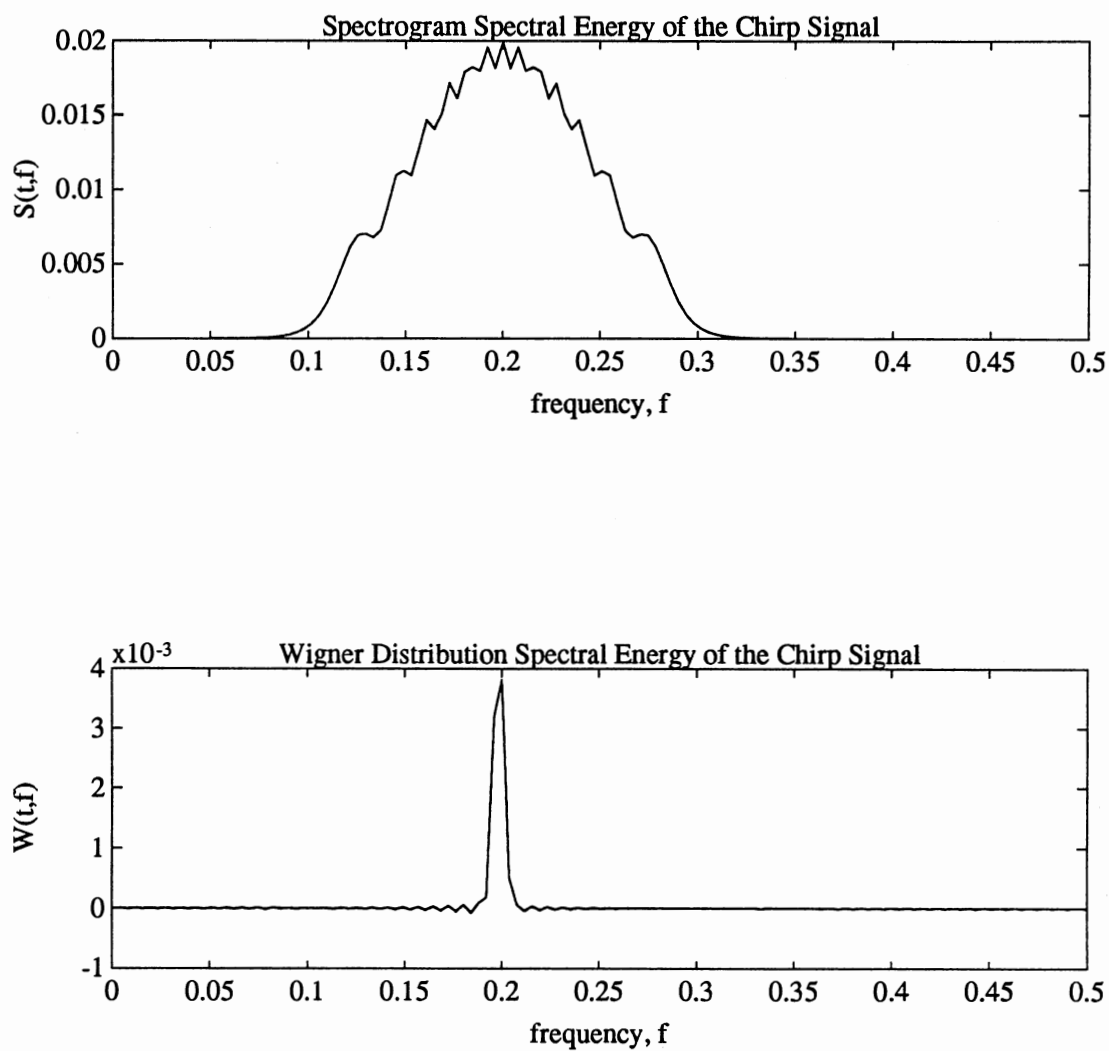


Figure III.5 Spectrogram and Wigner Distribution Spectral Energy Estimates of a Chirp Signal ($t = 0.5$ secs)

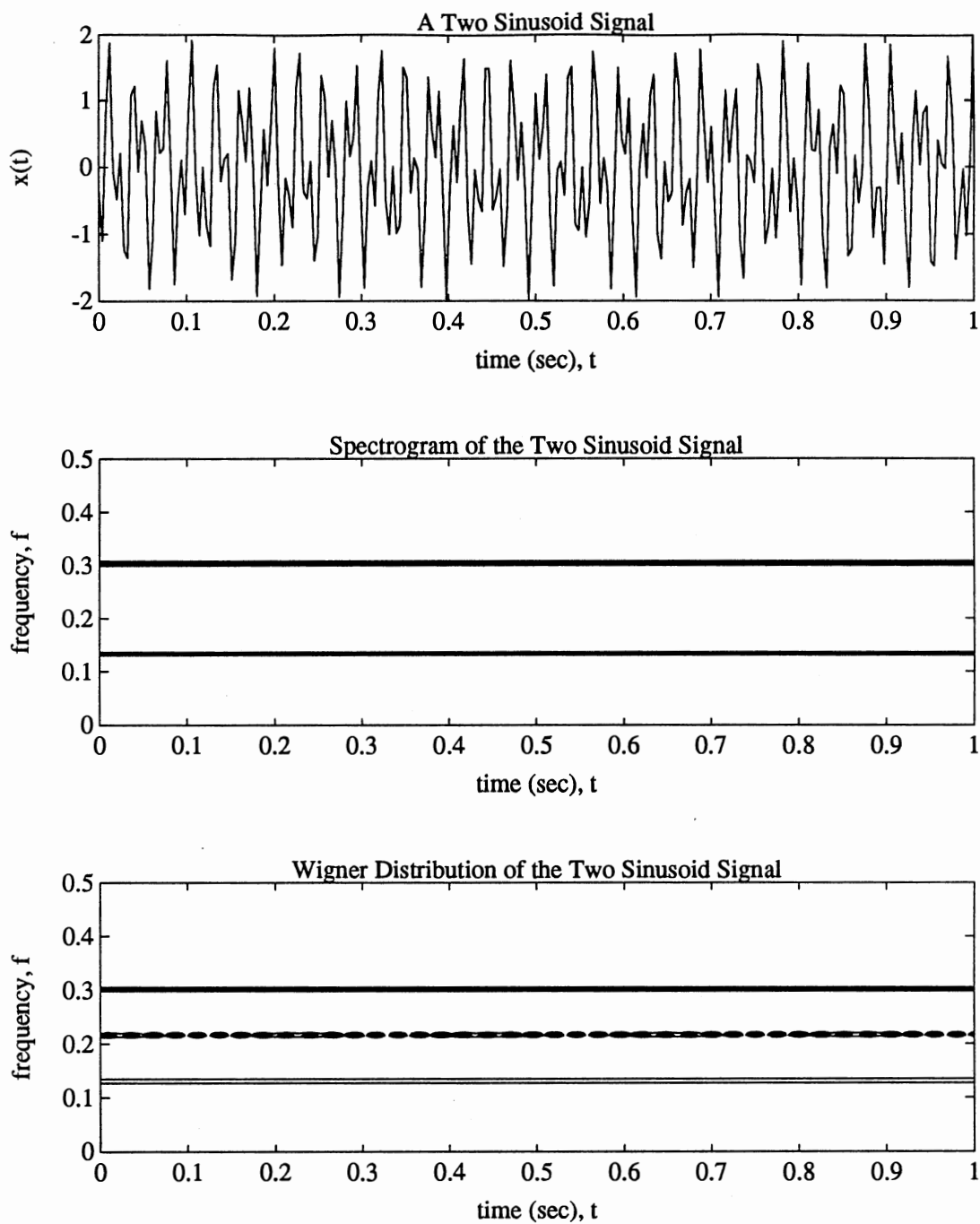


Figure III.6 The Spectrogram and Wigner distribution of Two Sinusoids

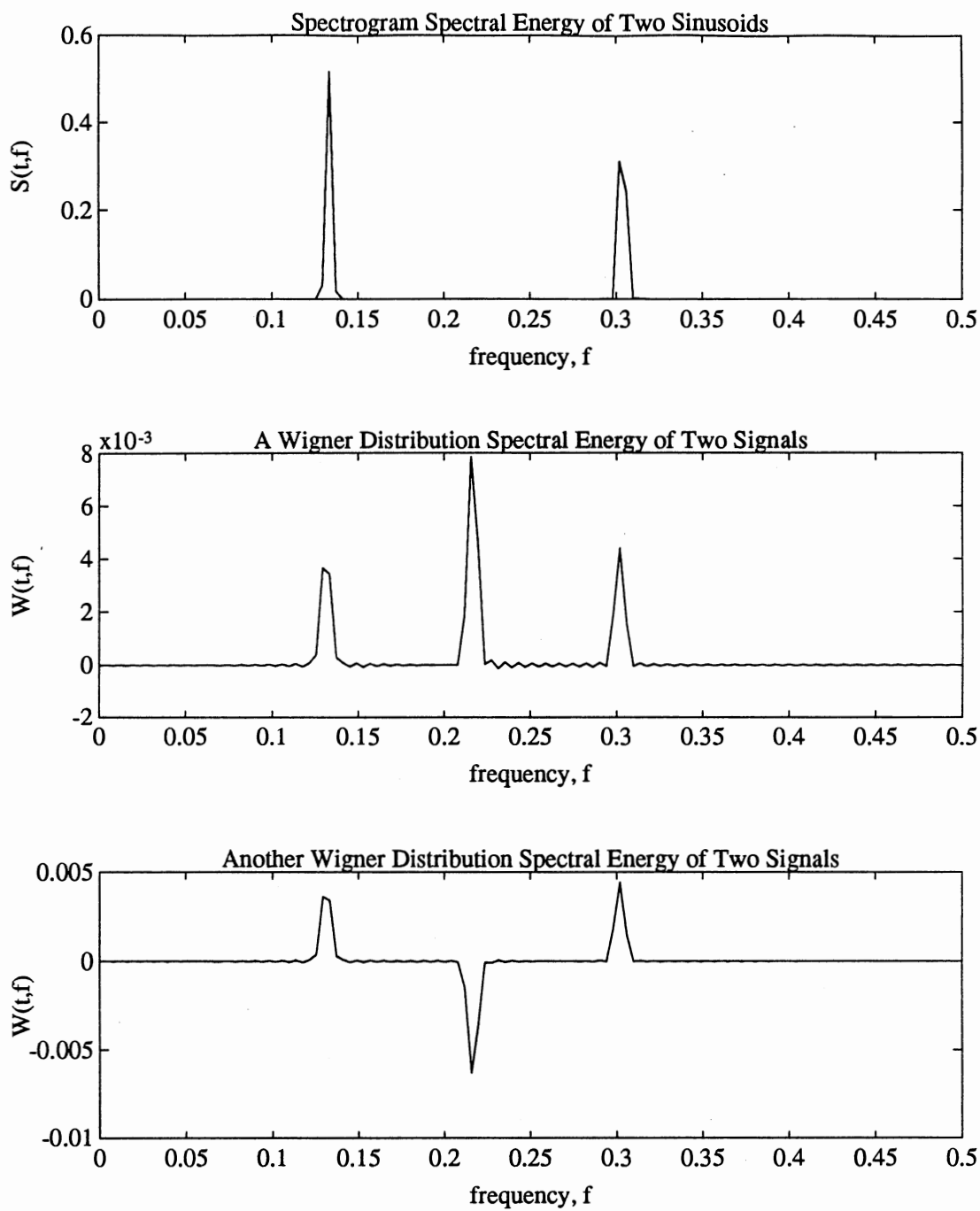


Figure III.7 Spectrogram and Wigner Distribution Spectral Energy Estimates of Two Sinusoids ($t = 0.5$ and 0.51 secs)

Crossterms result from the Wigner distribution acting on signals bilinearly. The bilinear nature of the Wigner distribution is not surprising since the ideal property

Property I.3: $x(t) + y(t) \Rightarrow TF_x(t, f) + TF_y(t, f)$

was missing from the Wigner distribution properties. In fact, the sum of any two signals produces the following undesirable Wigner distribution property:

Property III.12: $x(t) + y(t) = W_x(t, f) + W_y(t, f) + 2Re[W_{xy}(t, f)]$

where $W_{xy}(t, f)$ is the *cross Wigner distribution* defined by

$$W_{xy}(t, f) = \int_{-\infty}^{\infty} x(t + \tau/2) y^*(t - \tau/2) e^{-2\pi f \tau} d\tau, \quad (\text{III.14})$$

for any signals $x(t)$ and $y(t)$. Property III.12 indicates that for every two sinusoids, the Wigner distribution produces one crossterm, $2Re[W_{xy}(t, f)]$. During initial research into the Wigner distribution, some researchers believed crossterms contained additional signal information for time varying signals [Abe89]. However, researchers now agree that crossterm components have no physical significance and they are artifacts [Coh89] and [Cho89]. Thus, crossterms components present a negative aspect to the Wigner distribution.

Figure III.7 shows the spectral energy estimate produced by the spectrogram at $t = 0.5$ sec. Figure III.7 further shows two different spectral energy estimates obtained from the Wigner distribution at times $t = 0.5$ sec and $t = 0.51$ sec. The Wigner distribution spectral energy estimates bring out two points. First, the crossterm component is negative valued, and second the crossterm amplitude varies with time.

Negative Wigner distribution values defy the ideal time-frequency representation property of

Property I.2 and II.2: $TF_x(t, f) \geq 0$.

Furthermore, negative Wigner distribution values cause signal processing algorithms two main problems. First, spectral energy estimates based upon a squared modulus function that produces only non-negative values. Thus, difficulties arise when interpreting the Wigner distribution as a spectral energy. Second, non-negative values disallow interpreting the Wigner distribution as a two-dimensional pdf distributing energy over time and frequency. Thus, many probability based techniques are not directly applicable to the Wigner distribution.

Additional problems arise for the Wigner distribution because crossterm amplitudes change with time. For the most part, changing crossterm values disallow many methods for crossterm removal. Thus, signal processing techniques cannot remove crossterm components from the Wigner distribution. Overall, this example indicates that the Wigner distribution does not allow estimating spectral energies since the Wigner distribution spectral energy estimates contain negative values and crossterms. The next example shows the extent to which crossterms disturb Wigner distribution based spectral energy estimates.

The next example extends the last example by showing how the Wigner distribution deals with more than two sinusoids. This example's signal is produced by

$$x(t) = \sum_{i=1}^5 \sin(2\pi f_i t), \quad (\text{III.15})$$

where $f_1 = 2.6$ Hz, $f_2 = 19.3$ Hz, $f_3 = 33.9$ Hz, $f_4 = 77.3$ Hz, and $f_5 = 89.5$ Hz. Figure III.8 displays the five sinusoid signal, the spectrogram, and the Wigner distribution. Notice that the spectrogram contains five components while the Wigner distribution contains many more. The Wigner distribution crossterms are more clearly visible in Figure III.9

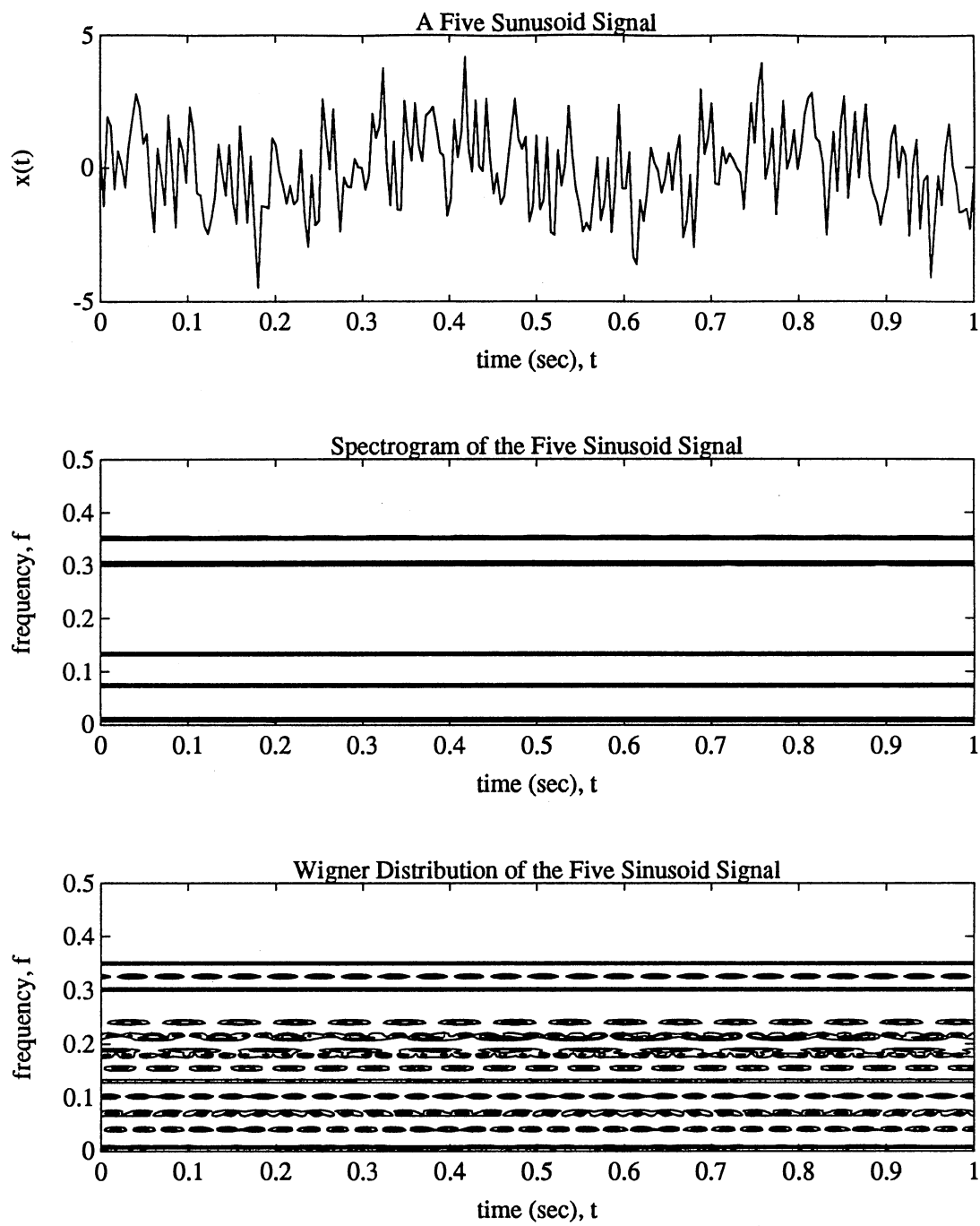


Figure III.8 The Spectrogram and Wigner distribution of Five Signals

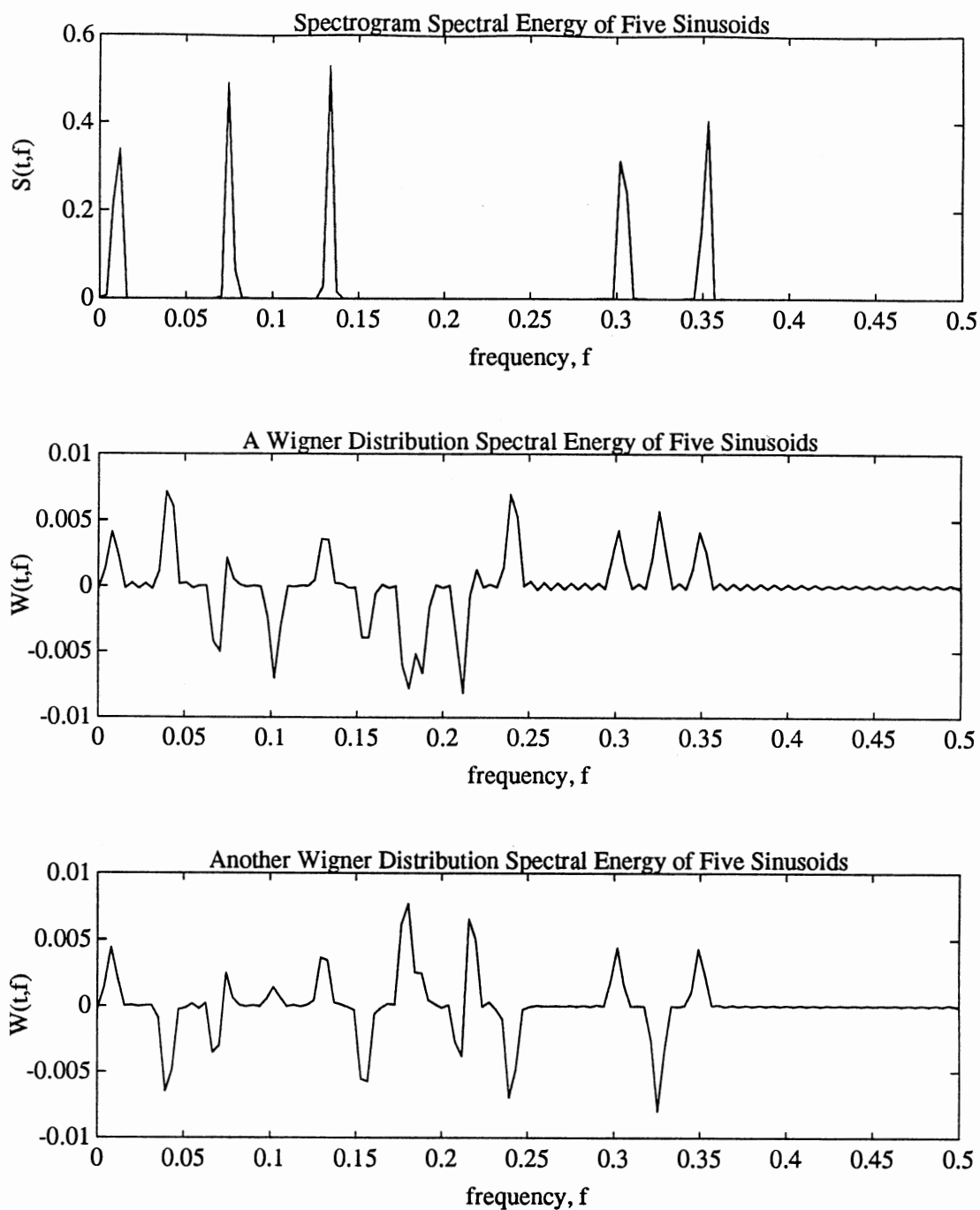


Figure III.9 Spectrogram and Wigner Distribution Spectral Energy Estimates of Five Signals ($t = 0.5$ and 0.51 secs)

which gives the spectrogram spectral energy estimate and two Wigner distribution spectral energy estimates. The Wigner distribution spectral energies indicate that crossterm components completely overwhelm the true spectral energies. Additionally, both Wigner distribution based spectral energy estimates appear quite different. However, the only difference between the two distribution spectral energies is the amplitudes of the cross-terms. Property 3.12 indicates the frequency of each crossterm. For a signal containing N sinusoids of frequency f_i where $i \in [1, N]$, there exist

$$\# \text{ crossterms} = \binom{N}{2} = \frac{N(N-1)}{2} \quad (\text{III.16})$$

each with frequency

$$f_{\text{crossterm},j} = \frac{f_j + f_k}{2}, \quad (\text{III.17})$$

where $j, k \in [1, N(N-1)/2]$ and j and k are chosen by combinations.

The five sinusoid signal example indicates that the Wigner distribution does not produce accurate spectral energy estimates due to crossterms components. Therefore, the Wigner distribution does not provide accurate spectral energy estimates for any broadband signal because broadband signals contain a large number of sinusoids. Because many signals are broadband, the Wigner distribution cannot replace the spectrogram. The question remains whether any time varying signals exist besides chirp signals where the Wigner distribution can replace the spectrogram. One possible signal is the organ signal from Chapter I. The organ signal represents the concatenation of single sinusoids. Since only one sinusoid occurs for any one time, the Wigner distribution may more accurately indicate when a frequency change occurs. An example organ signal, defined as

$$x(t) = \sin(2\pi ft) \quad (\text{III.18})$$

where

$$f = \begin{cases} 4.20\text{Hz} & \forall t \in [0.0, 0.33] \\ 48.4\text{Hz} & \forall t \in (0.33, 0.67] \\ 13.5\text{Hz} & \forall t \in (0.67, 1.0] \end{cases}, \quad (\text{III.19})$$

is presented in Figure III.10 along with the spectrogram and the Wigner distribution. Figure III.11 shows the spectrogram spectral energy estimate and the Wigner distribution spectral estimate for $t = 0.33$ secs. Notice that detecting the first frequency change at $t = 0.33$ is not feasible using the spectrogram. The spectrogram spreading has smeared away a detectable transition region. The Wigner distribution contour indicates where the transition region occurs. Comparing the spectrogram and Wigner distribution spectral energies at $t = 0.33$ indicates that the spectrogram estimates that both sinusoids occur simultaneously while the Wigner distribution produces noise. The Wigner distribution noise occurs since the only major positive Wigner coefficient appears at the crossterm frequency. Thus for the examples presented in this chapter, the Wigner distribution can only replace the spectrogram for chirp signals.

Since the Wigner distribution does not provide accurate spectral energy estimates for most signals, how is the Wigner distribution useful in signal processing? In the next section, examples show that functions computed on the Wigner distribution may provide useful time varying signal information. Chapter V illustrates an example signal processing system that relies on these functions to determine whether a signal is stationary or nonstationary. Thus, the Wigner distribution, although theoretically important, does not always solve one of the main goals of this thesis: the attempt to replace the spectrogram with a time-frequency representation.

The Discrete Wigner Distribution

To this point, all time-frequency representations, including the Wigner distribution, have been discussed in context of continuous variables. Since most signal processing algorithms use digital signals, implementing the Wigner distribution requires defining a discrete Wigner distribution. Although discretizing the Wigner distribution appears simple, 33 years lapsed from the time Ville applied the Wigner distribution until a discrete Wigner distribution appeared. Classen and Mecklenbrauker, [Cla80b], define the discrete

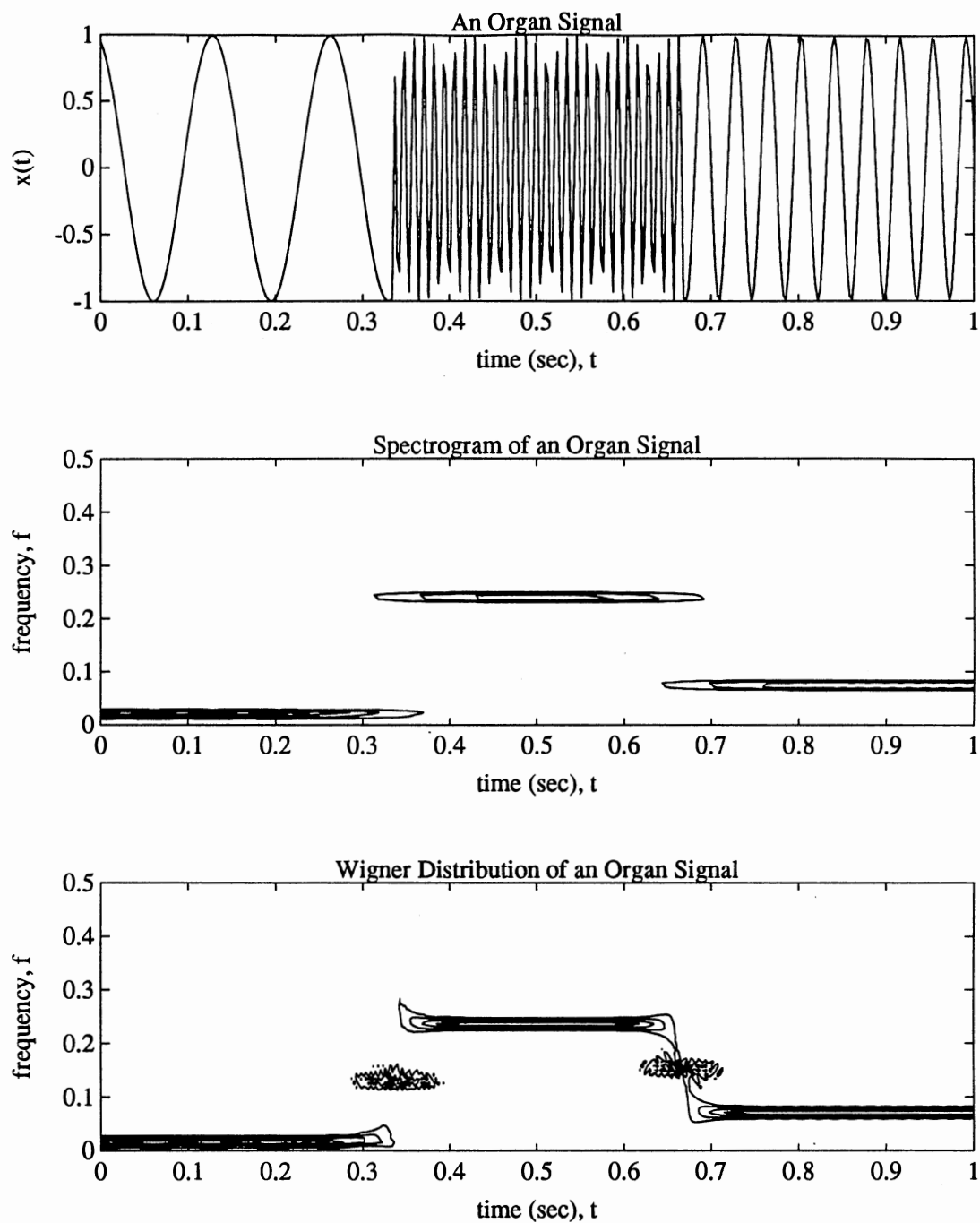


Figure III.10 The Organ Signal, the Corresponding Spectrogram, and the Corresponding Wigner Distribution

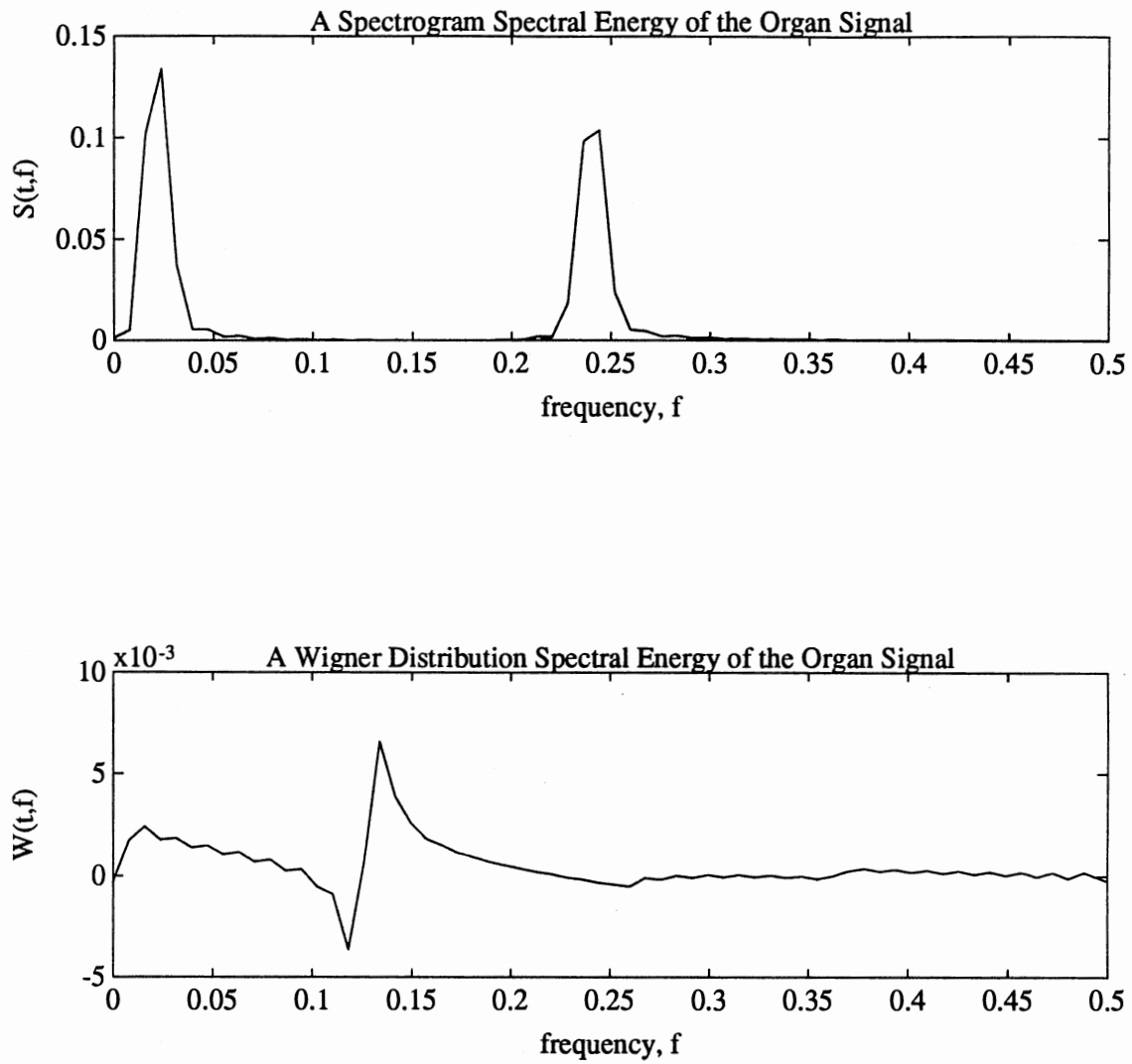


Figure III.11 Spectrogram and Wigner Distribution Spectral Energy Estimates of an Organ Signal ($t = 0.5$ secs)

Wigner distribution as

$$W_x(n, \theta) = \sum_{k=-\infty}^{\infty} x(n+k)x^*(n-k)e^{-j2k\theta} \quad (\text{III.20})$$

where $n \in Z$ implies that n is an integer and $\theta \in [0, 2\pi]$ is a continuous radian frequency.

Defining a discrete Wigner distribution is difficult because the 2 in the $\tau/2$ term of the continuous Wigner distribution (III.2) does not permit a usual discretization in terms of τ . The discretization of τ is difficult since for $\tau \in (2Z^+ + 1)$ where Z^+ represents the positive integers (meaning the τ is an odd integer), $\tau/2 \notin Z$ since $\tau/2$ produces a real number with a non-zero fractional part. Classen and Mecklenbrauker resolved the $\tau/2$ problem by constraining τ so that $t = n \in Z$ and

$$W_x(n, f) = \sum_{\tau=-\infty}^{\infty} x(n + \tau/2)x^*(t - \tau/2)e^{-i\tau 2\pi f}. \quad (\text{III.21})$$

Substituting $k = \tau/2$ produces

$$W_x(n, f) = \sum_{k=-\infty}^{\infty} x(n+k)x^*(n-k)e^{-i(2k)(2\pi f)}, \quad (\text{III.22})$$

which equals (III.20). This discretization presents a problem, however, since the $e^{-i2k\theta}$ term of (III.20) should be $e^{-ik\theta}$. As such $e^{-ik\theta}$ is periodic with 2π , whereas $e^{-2k\theta}$ is periodic with π . Thus, if (II.22) was implemented in a signal processing system, then the Wigner distribution would contain incorrect values if the sampling rate was less than twice the Nyquist rate. Flawed results occur since $e^{-2k\theta}$, being periodic in π , implies frequencies from π to 2π wrap back around (*aliasing*) the frequency domain [Cla83]. Discrete Wigner distribution aliasing prohibits using the discrete Wigner distribution directly on signals. However, two methods exist which circumvent the aliasing problems.

Aliasing Problems of the Discrete Wigner Distribution

Because the discrete Wigner distribution is periodic in π , aliasing occurs when a system samples a signal at less than four times the signal's bandwidth. Thus, if the system sampling rate exceeds four times the signal's bandwidth (or twice the Nyquist rate), then discrete Wigner distribution aliasing produces no side-effects. Requiring the higher sampling rate is feasible for some applications. However, many signals already exist from systems designed using the normal Nyquist rate. Thus, raising the minimum sampling rate does not totally eliminate all discrete Wigner distribution aliasing problems.

The other method for eliminating discrete Wigner distribution aliasing requires computing the analytic signal pair. An *analytic signal* $s(t)$ of $x(t)$ is defined as

$$s(t) = x(t) + ix\hat{(t)} \quad (\text{III.23})$$

where

$$x\hat{(t)} = \text{CPV} \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\xi)}{t - \xi} d\xi. \quad (\text{III.24})$$

The $x\hat{(t)}$ term of (III.23) and (III.24) is called the *Hilbert transformation* of $x(t)$. For discrete $s(n)$, the analytic signal is obtained by

$$S(f) = \begin{cases} 2X(f) & \forall f > 0 \\ X(f) & \forall f = 0, \\ 0 & \forall f < 0 \end{cases} \quad (\text{III.25})$$

where $X(f)$ is the Fourier transformations of $x(t)$. Because $S(f)$ contain non-zero values for $f \geq 0$, $S(f)$ has half the bandwidth than $X(f)$. Since the analytic signal, $s(t)$, contains half the bandwidth of the original signal $x(t)$, the Nyquist rate for $s(t)$ is half of $x(t)$'s rate. Therefore, the discrete Wigner distribution of $s(t)$ produces no aliasing problems.

One positive aspect of using the analytic signal is that low frequency crossterm component amplitudes reduce when using the analytic signal compared to the actual signal [Boa87], [Boa88], and [Fla84]. One negative aspect of analytic signals is that

analytic signals are always complex valued. However, the discrete Wigner distribution is computed using the discrete Fourier transformation. Since the Fourier transformation accepts complex valued signals, the complex analytic signal values are not a problem.

Implementing the Discrete Wigner Distribution

Because only the t term in (III.20) is discretely valued, the frequency variable requires discretizing. The complete discrete Wigner distribution is

$$W_x(n, m) = \sum_{k=-N}^N x(n+k)x^*(n-k)e^{-i2\pi km/2N}, \quad (\text{III.26})$$

where n represents discrete time, m represents discrete frequency, and the number of samples in x is $(2N + 1)$, an odd number. For perfect frequency resolution, N should approach infinity. However, N approaching infinity requires an infinite number of samples. Thus, the number of samples is reduced by windowing $x(n)$. A windowed discrete Wigner distribution is called the *pseudo-Wigner distribution* and is defined by

$$W_x(n, m) = \sum_{k=-N}^N h(k)x(n+k)h^*(-k)x^*(n-k)e^{-i2\pi km/2N}, \quad (\text{III.27})$$

where $h(k)$ is a window function of length $(2N + 1)$. At first glance, windowing would seem to reduce the Wigner distribution's effectiveness on time varying signals. However, windowing the discrete Wigner distribution only reduces frequency resolution and does not effect the time resolution [Cla80b]. Thus, windowing reduces computation time while not producing any serious negative side effects.

One problem with the pseudo-Wigner distribution is that implementing (III.27) requires computing a discrete Fourier transformation of length $(2N + 1)$. Because $(2N + 1)$ is always odd, the conventional fast Fourier transform, which is based on lengths of 2^n , cannot compute the pseudo-Wigner distribution. Day and Yarlalagadda [Day87] resolved this problem by defining a new pseudo-Wigner distribution as

$$W_x(n, m) = \sum_{k=-N}^{N-1} h(k+1)x(n+k+1)h(-k)x(n-k)e^{-i2\pi(k+1/2)(m+1/2)/2N}, \quad (\text{III.28})$$

which allows using the fast Fourier transformation. Day showed that this definition maintains most of the properties found with the usual pseudo-Wigner distribution.

For most Wigner distribution implementations, (III.28) computes the discrete Wigner distribution. Problems associated with the odd sequence length are eliminated by appending a zero, called *zero padding*, on the end of the signal sequence. Since zero padding changes only the spectral resolution of a signal, the Wigner distribution information does not change.

Wigner Distribution Based Features

As previously discussed, the Wigner distribution is heavily researched because the Wigner distribution can generate some of the other time-frequency representations and because the Wigner distribution may contain time varying signal information. Thus, understanding the Wigner distribution allows understanding of other time-frequency representations. Recall that the Wigner distribution was constrained to having useful marginal distributions.

The idea is to generate signals from the Wigner distribution that contain information about a time varying signal. These signals are called *features* since the signals describe some signal characteristic or signal feature. Wigner distribution based features can then be used in signal processing algorithms for a variety of applications. These applications include detecting the presence of signals, detecting spectral changes, and monitoring signal power. Although the features are not as useful as a spectral energy estimate in some situations, the features can improve signal processing systems that contain time varying signals, especially pattern recognition systems. The next section describes and illustrates some of the main features generated from the Wigner distribution.

Basic Distribution Features

The first two Wigner distribution features arise directly from the original desired properties (Properties III.1 and III.2) of the Wigner distribution. Thus,

$$|x(t)|^2 = \int_{-\infty}^{\infty} W_x(t, f) df \quad (\text{III.29})$$

and

$$|X(f)|^2 = \int_{-\infty}^{\infty} W_x(t, f) dt, \quad (\text{III.30})$$

the instantaneous power and the spectral energy, provide the first two potential features. Additional features can be defined by applying different functions to the Wigner distribution. However, one approach, in particular, produces good results. The approach comes from the two-dimensional probability density functions (pdfs) time-frequency representation ideal properties. Although the Wigner distribution contains negative values, the Wigner distribution can be made non-negative simply by using the absolute value of the Wigner distribution. However, using the Wigner distribution's absolute value disallows using some of the Wigner distribution's properties. Thus, two-dimensional pdf theory can be applied to the Wigner distribution either by creating a non-negative Wigner distribution or by realizing that results obtained for the Wigner distribution values do not correspond directly to two-dimensional pdf based signals. Recall that if $W_x(t, f)$ is considered a two-dimensional pdf of signal energy, then two general moments,

$$m_k(t) = \int_{-\infty}^{\infty} \xi^k W_x(t, \xi) d\xi \quad (\text{III.31})$$

and

$$n_k(f) = \int_{-\infty}^{\infty} \xi^k W_x(\xi, f) d\xi \quad (\text{III.32})$$

can be defined. Since $W_x(t, f)$ is not an actual pdf, moments computed on the Wigner distribution do not follow the appropriate probability properties; however, these moments still may contain time varying spectral energy information.

Using the two-dimensional pdf based moments method, notice that $k = 0$ produces

$$m_0(t) = \int_{-\infty}^{\infty} W_x(t, \xi) d\xi \quad (\text{III.33})$$

and

$$n_0(f) = \int_{-\infty}^{\infty} W_x(\xi, f) d\xi, \quad (\text{III.34})$$

which equal the instantaneous power and the spectral energy, respectively. Since $m_0(t)$ is the marginal distribution of $W_x(t, f)$ with respect to t , $m_0(t)$ represents the distribution of power with respect to time. Similarly, $n_0(f)$ (the spectral energy) indicates how energy is distributed over frequency. Therefore, locations where $m_0(t)$ and $n_0(f)$ contain large values correspond to areas in the time-frequency domain where signal energy occurs. Thus, the pdf approach of generating Wigner distribution features produces an intuitive meaning for the first two moments.

The second set of features occurs by setting $k = 1$ and scaling by $m_0(t)$ and $n_0(f)$.

The two moments defined with $k = 1$ are

$$m_1(t) = \frac{1}{m_0(t)} \int_{-\infty}^{\infty} \xi W_x(t, \xi) d\xi \quad (\text{III.35})$$

and

$$n_1(f) = \frac{1}{n_0(f)} \int_{-\infty}^{\infty} \xi W_x(\xi, f) d\xi \quad (\text{III.36})$$

where $m_1(t)$ is called the instantaneous frequency, $n_1(f)$ is called the group delay, and both $m_1(t)$ and $n_1(f)$ are defined only when $m_0(t)$ and $n_0(f)$ do not equal zero, respectively. Both the instantaneous frequency and the group delay relate directly to the mean values

of $W_x(t, f)$ for the appropriate variables. Thus, the instantaneous frequency represents the mean frequency for a particular time while the group delay states the average time a particular frequency occurs. The reader should note that the instantaneous frequency and the group delay existed in signal processing long before Ville applied the Wigner distribution to signal processing.

Throughout the current literature, no other moment based features have been defined. Thus, the instantaneous power, the instantaneous frequency, the spectral energy, and the group delay comprise the basic Wigner distribution features. The following three examples indicate how these four features perform on both the Wigner distribution and the spectrogram. The first example uses the chirp signal given by (I.11). Figure III.12 illustrates the chirp signal, the spectrogram instantaneous power function, and the Wigner distribution instantaneous power function. Notice that the Wigner distribution instantaneous power seems noisier than the spectrogram instantaneous power. This may indicate that the Wigner distribution tracks signal energy changes better than the spectrogram. Figure III.13 again illustrates the chirp signal but with the spectrogram instantaneous frequency and the Wigner distribution instantaneous frequency. Note that these features can be compared directly to the signal since all time scales match. Both the spectrogram and the Wigner distribution instantaneous frequency track the linearly increasing frequency of the chirp signal. Thus, both time-frequency representations produce good time varying spectral energy information. As for whether the Wigner distribution instantaneous power is better than the spectrogram instantaneous power, no decision can be made without additional analysis.

Although the frequency based features are not used much in practice, they are displayed for the chirp signal just to see how they respond. Figure III.14 displays a the spectrogram power spectral density and the Wigner distribution power spectral density.

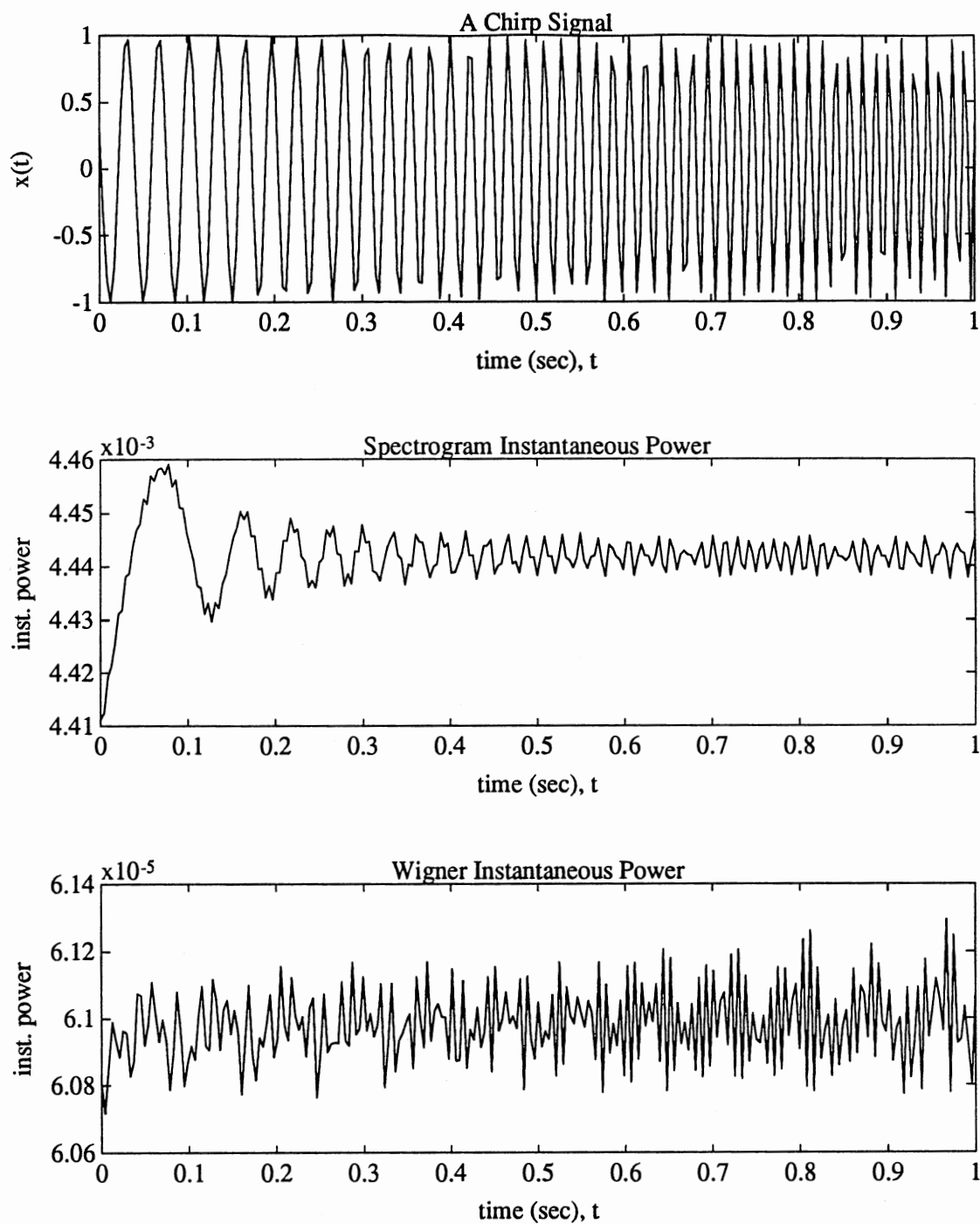


Figure III.12 The Spectrogram Instantaneous Power and the Wigner Distribution Instantaneous Power of a Chirp Signal

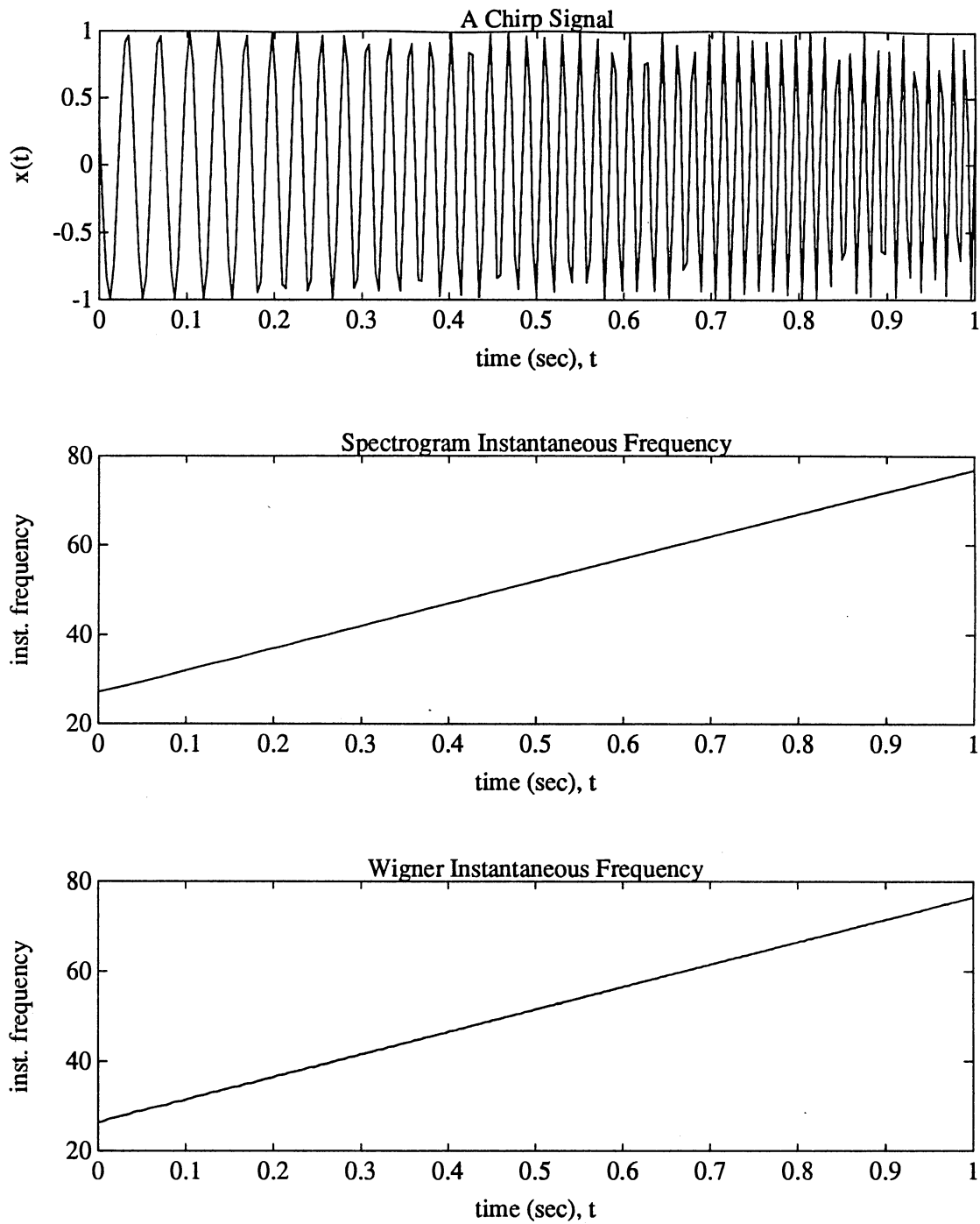


Figure III.13 The Spectrogram Instantaneous Frequency and the Wigner Distribution Instantaneous Frequency of a Chirp Signal

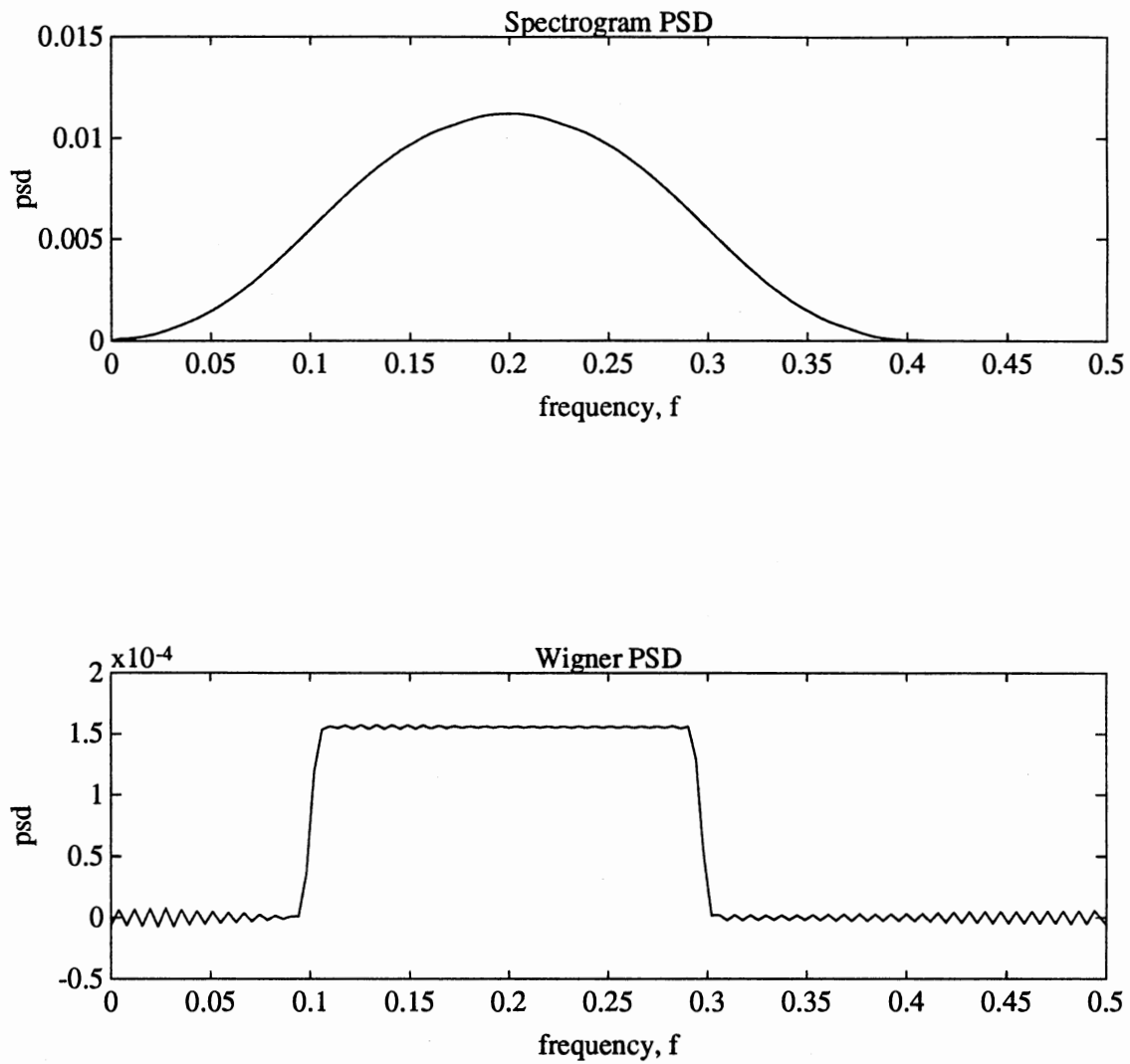


Figure III.14 The Spectrogram Power Spectral Density and the Wigner Distribution Power Spectral Density for a Chirp Signal

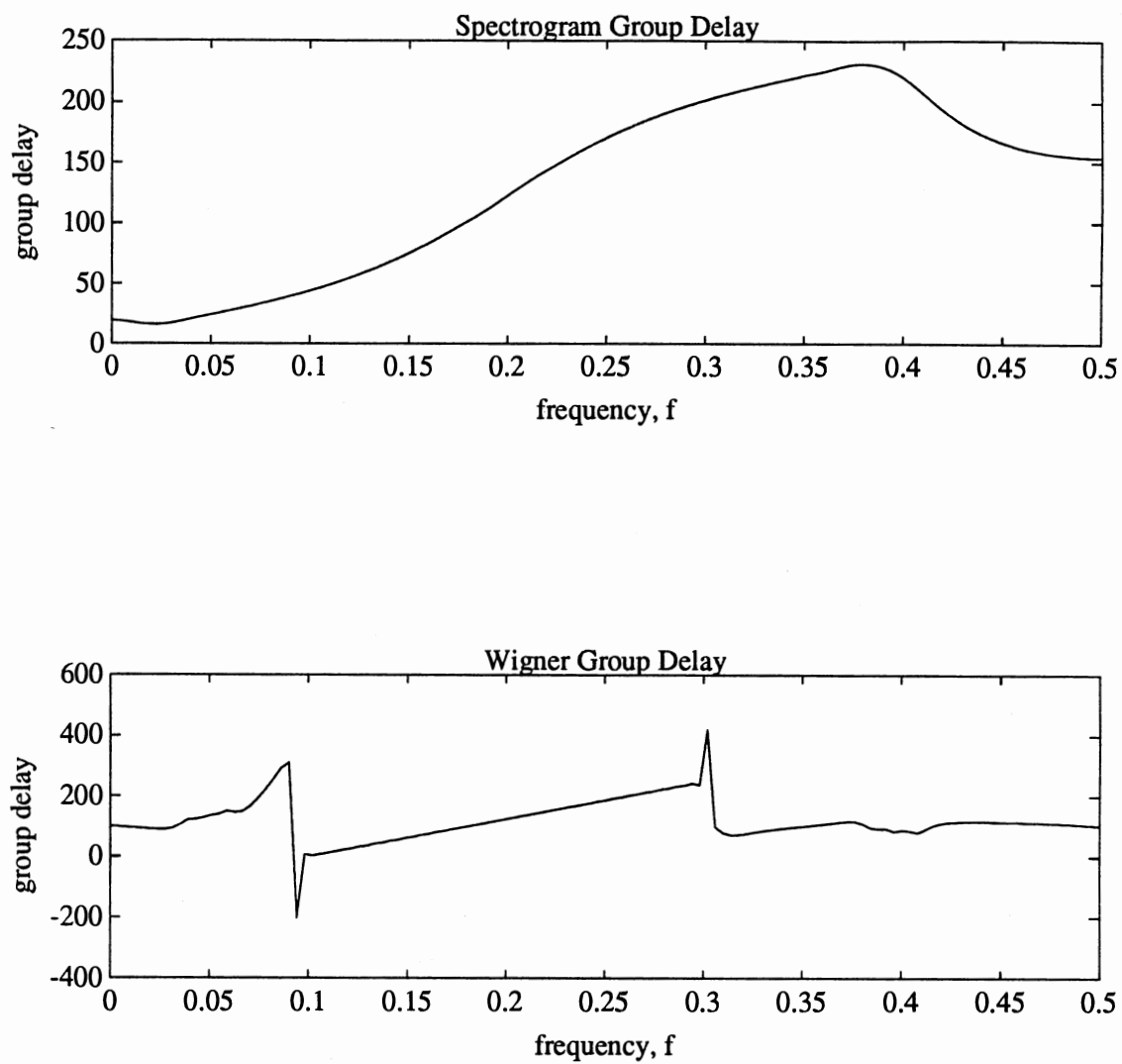


Figure III.15 The Spectrogram Group Delay and the Wigner Distribution Group Delay for a Chirp Signal

Notice how the spectrogram power spectral density shows signs of the Gaussian window while the Wigner distribution power spectral density does not. This indicates that the Wigner distribution may provide additional information during transition regions over the spectrogram. Additionally, notice how the Wigner distribution power spectral density more accurately predicts the starting and stopping frequencies of the chirp signal. Figure III.15 illustrates graphs of the spectrogram group delay and the Wigner distribution group delay. As with other features, the Wigner distribution group delay appears less smooth than the spectrogram group delay. Additionally, the starting and stopping frequencies of the chirp signal are certainly noticeable in the Wigner distribution group delay and not in the spectrogram group delay. Another interesting point to make about Figure III.15 is that some of the Wigner distribution group delay values are negative while the spectrogram group delay values are all non-negative. Thus, the fact that the Wigner distribution produces negative values is evident from the example. The overall conclusion from the frequency dependent features is that the Wigner distribution seems to provide faster responses since the starting and stopping frequencies of the chirp signal are more evident from the Wigner distribution than from the spectrogram.

The second example illustrates the example organ melody introduced in (I.15). Figure III.16 displays an organ melody, the spectrogram instantaneous power, and the Wigner distribution instantaneous power. Once again, the Wigner distribution instantaneous power seems to change values more rapidly than the spectrogram instantaneous power. However, this example indicates that the Wigner distribution may provide an additional feature generation capability because the Wigner distribution instantaneous power contains quickly changing values at locations where the organ melody changes frequency. The spectrogram instantaneous power does not indicate these changing frequencies and the transition regions as clearly when compared to the Wigner distribution instantaneous power. Figure III.17 illustrates the organ melody again but with the spectrogram instantaneous frequency and the Wigner distribution instantaneous

frequency. Notice how quickly the Wigner distribution instantaneous frequency tracks changes in frequency when compared to the spectrogram instantaneous frequency. Additionally, estimating the single frequency for a particular time is not possible for much of the spectrogram instantaneous frequency; whereas, the Wigner distribution instantaneous frequency provides much better estimates. Thus, the organ melody example seems to indicate that the Wigner distribution produces better time varying spectral features when compared to the spectrogram features. Noting from the previous discussion, a single example will not illustrate all good and bad aspects of the Wigner distribution and therefore consider the third example which investigates how features appear for broadband signals.

The third example uses the speech signal discussed in Chapter I. This speech signal represents a male speaking the first four letters of the word SIGNAL. Figure III.18 provides a speech signal, the spectrogram instantaneous power, and the Wigner distribution instantaneous power. Notice that the speech signal contains changing characteristics over time. As was found with the two previous examples, the Wigner distribution instantaneous power seems to contain more time varying spectral information over the spectrogram instantaneous power. The additional information results from the Wigner distribution instantaneous power changing more rapidly compared to the spectrogram instantaneous power. Additionally, notice how large Wigner distribution instantaneous power magnitudes correspond to the large magnitudes in the speech signal. Figure III.19 illustrates the speech signal again but with the spectrogram instantaneous frequency and the Wigner distribution instantaneous frequency. Again, the Wigner distribution instantaneous frequency changes more rapidly than the spectrogram instantaneous frequency. In fact, the Wigner distribution instantaneous frequency appears quite similar to the Wigner distribution instantaneous power whereas the spectrogram instantaneous frequency differs significantly from the spectrogram instantaneous power.

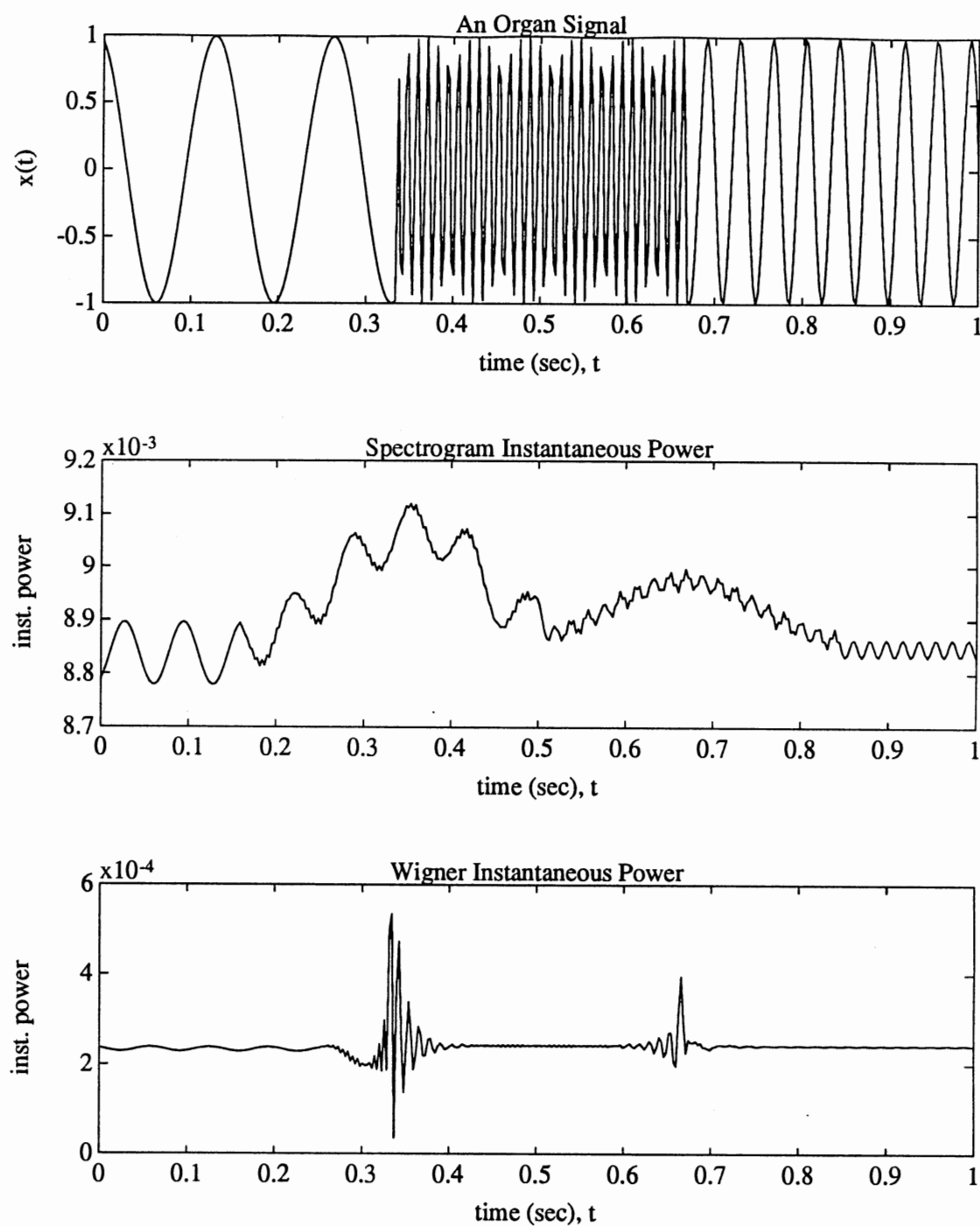


Figure III.16 The Spectrogram Instantaneous Power and the Wigner Distribution Instantaneous Power of an Organ Melody Signal

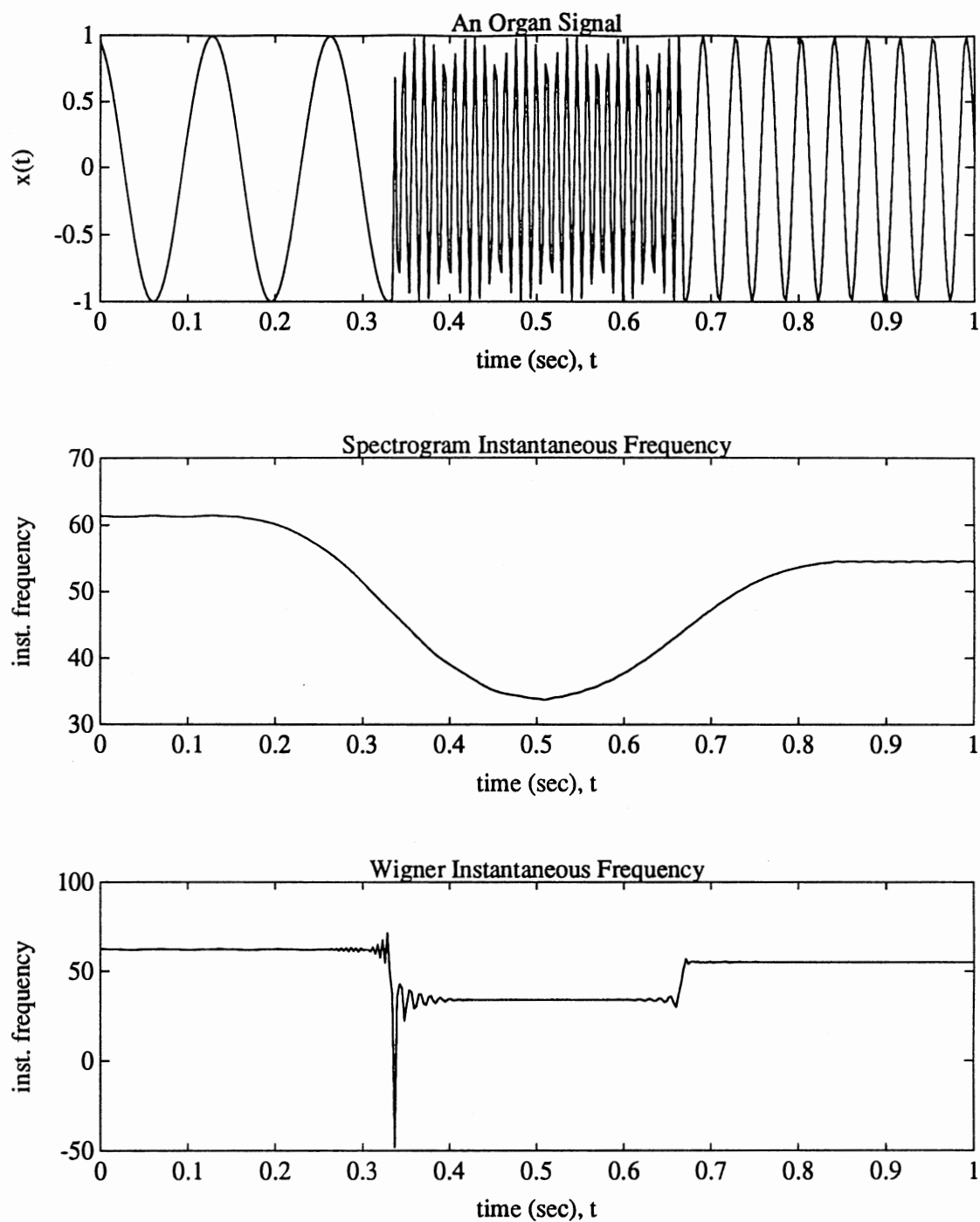


Figure III.17 The Spectrogram Instantaneous Frequency and the Wigner Distribution Instantaneous Frequency of an Organ Melody Signal

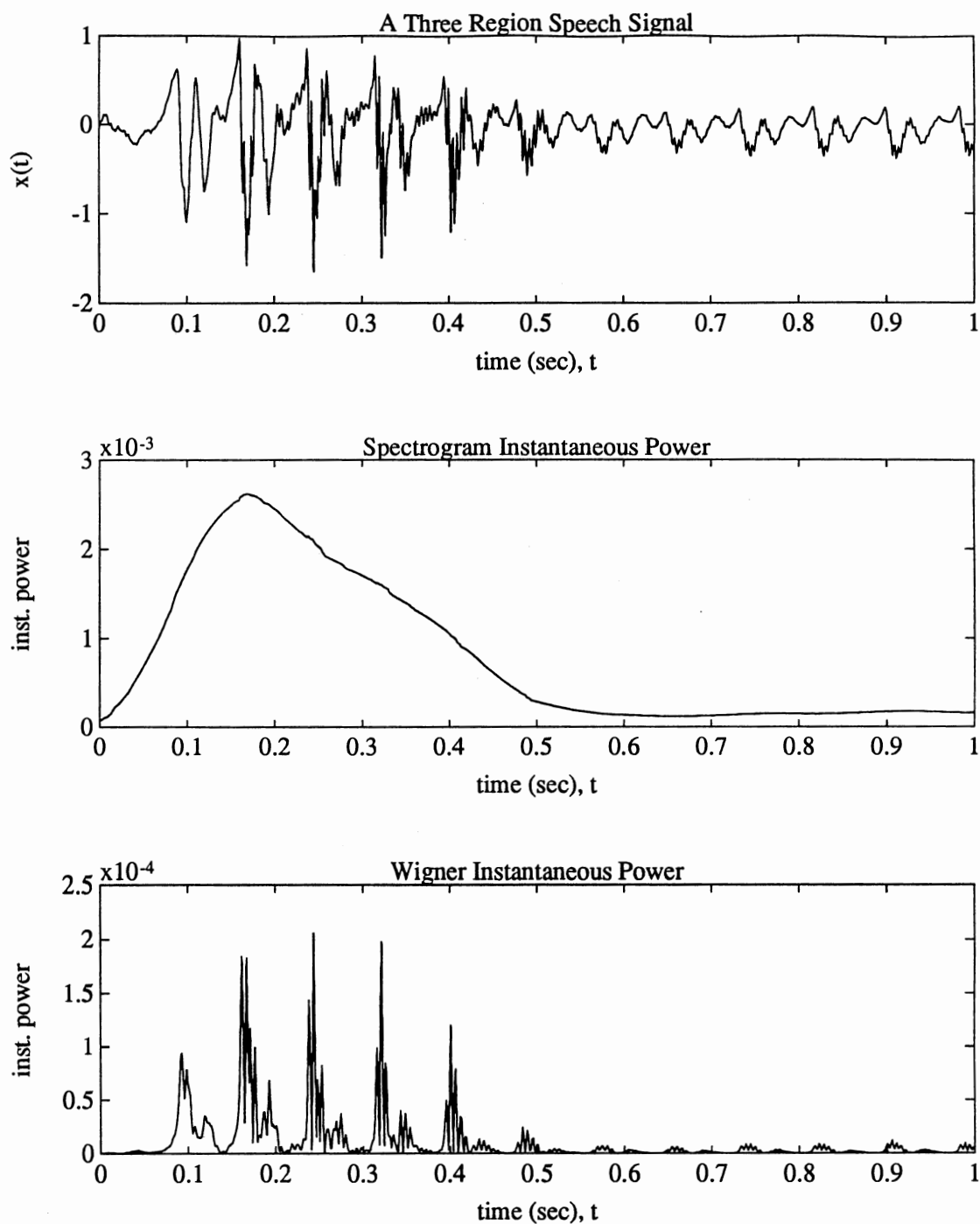


Figure III.18 The Spectrogram Instantaneous Power and the Wigner Distribution Instantaneous Power of a Speech Signal

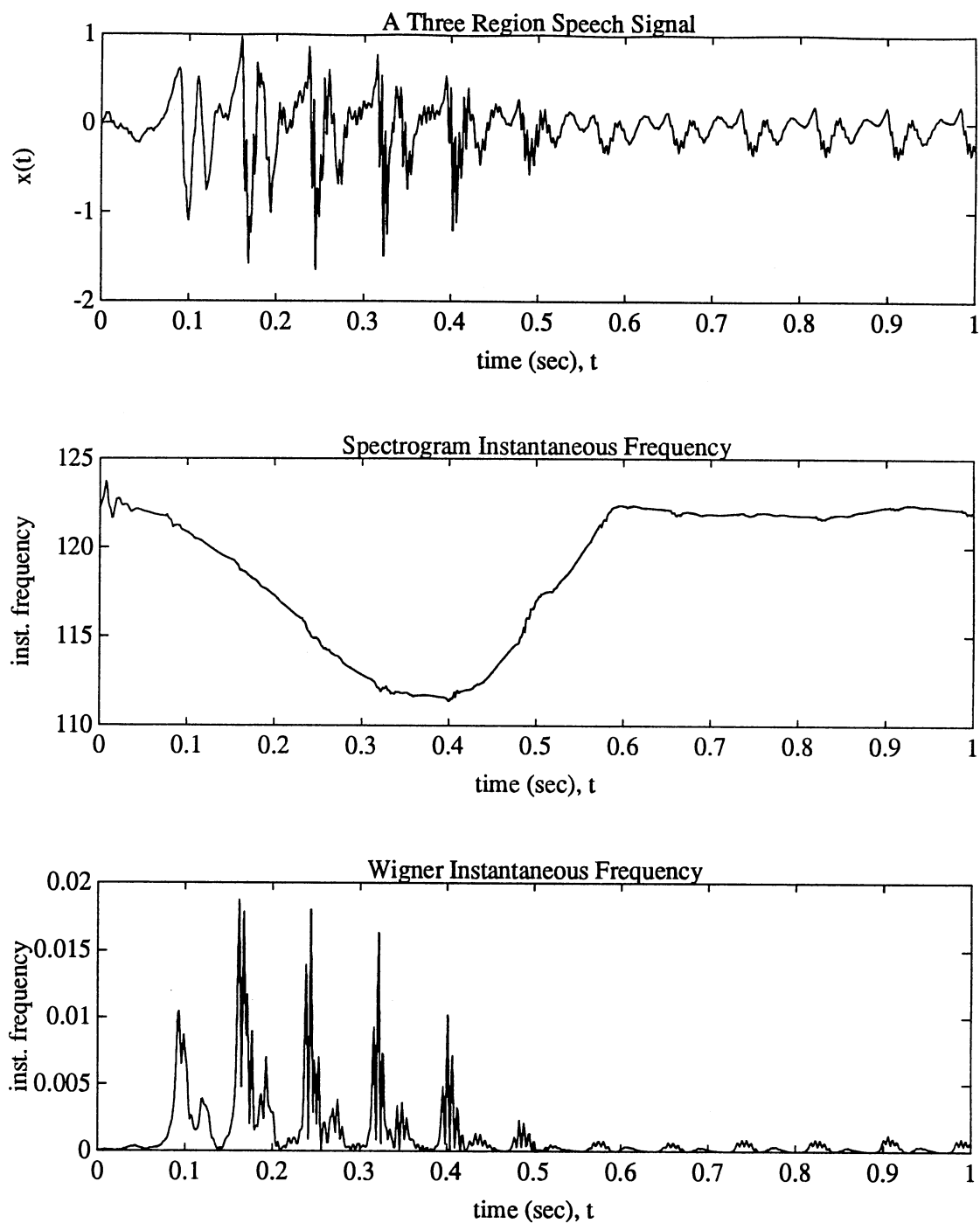


Figure III.19 The Spectrogram Instantaneous Frequency and the Wigner Distribution Instantaneous Frequency of a Speech Signal

The examples indicate that Wigner distribution features may provide additional information when compared to spectrogram features. Chapter V introduces a signal processing system that uses the two time dependent Wigner distribution features to identify transition regions in time varying signals. Before covering the transition region system, Chapter IV provides new information which substantiates the claim that the Wigner distribution, and in fact most time-frequency representations, are not capable of replacing the spectrogram for estimating time varying spectral energies. The next section describes some of the current Wigner distribution applications.

Signal Processing Applications of the Wigner Distribution

The Wigner distribution is currently being investigated by many researchers for use in signal processing algorithms [Coh89]. For instance, researchers are interested in locating time varying features for use in pattern recognition systems. Kumar and Carroll [Kum84] are using the Wigner distribution for detecting the presence and the beginning of target signals. Usui and Araki [Usu90] use the Wigner distribution to analyze body surface potential mappings (BPSM) that are a series of electrocardiograms (ECGs) at various positions on a body surface. Usui and Araki found that the Wigner distribution assists in removing redundant leads. Yen, Dragonette, and Numrich [Yen90] found that the Wigner distribution produces useful pattern recognition features for analyzing acoustic scattering from elastic objects.

Also being investigated are applications of transient signal detection. O'Shea and Boashash [Osh90] have found the instantaneous frequency derived from the Wigner distribution useful in detecting transient signals. Malkoff and Cohen [Mal90] use a neural network along with several features derived from the Wigner distribution to detect and to classify transient signals and signals embedded in noise.

Other interesting applications of the Wigner distribution have been reported. Chester, Taylor, and Doyle [Che83] found that the Wigner distribution was quite useful in several speech processing applications. Riley [Ril89] successfully located formant frequencies in vowels using gradient type algorithms on the Wigner distribution. Pries, Halwatsch, Bloom, and Deer [Pre87] use the Wigner distribution as a display tool for analyzing phase distortion in signal filters. Eggermont and Smith [Egg90] characterize auditory neurons using the Wigner distribution.

Chapter Summary

This chapter has reviewed both theoretical and practical aspects of the Wigner distribution. Even though the Wigner distribution was developed for quantum mechanics, Ville found that the Wigner distribution is quite useful in signal processing. The Wigner distribution usefulness stems directly from the three properties which constrained the Wigner distribution in the first place. Since these three properties (instantaneous power, spectral energy, and total energy) indicate that the Wigner distribution may provide useful time varying spectral estimates, the spectrogram and Wigner distribution spectral energy estimates were compared for several signals. Unfortunately, the Wigner distribution based spectral energy estimates contain crossterm components and negative values which render the Wigner distribution of little use as a time varying spectral energy estimator. However, by considering the Wigner distribution as a two-dimensional pdf, probability based features computed upon the Wigner distribution can be defined. Four of these features were computed for the spectrogram and the Wigner distribution on some example signals. The feature examples indicate that the Wigner distribution features may provide more time varying information than the spectrogram features. Wigner distribution features are investigated more thoroughly in Chapter V.

Since this chapter indicates that the Wigner distribution cannot replace the spectrogram as a time varying spectral energy estimator, does a time-frequency representation exist that can replace the spectrogram? Chapter IV investigates this question with a technique called the correlation approach to time-frequency representations. The correlation approach allows comparing most time-frequency representations, in terms of transforming variables of the nonstationary autocorrelation function. As such, many interesting relationships and properties concerning time-frequency representations will be analyzed.

CHAPTER IV

A CORRELATION APPROACH TO TIME-FREQUENCY REPRESENTATIONS

Introduction

This thesis has reviewed many aspects of time varying spectral energy analysis. Specifically, Chapter I reviewed basic signal processing terminology and presented examples of time varying signals. Chapter I also introduced two signal processing systems that require some form of time varying spectral energy information to perform correctly. Chapter II reviewed many of the techniques developed over the years that attempt estimating time varying spectral energies. Of the reviewed techniques, the spectrogram produces the most usable time varying spectral energy representation, even though the spectrogram assumes short-time time-invariant signals. For truly time varying signals, the other reviewed techniques (collectively named alternate time-frequency representations) generally cannot replace the spectrogram as an estimator of spectral energy. However, Chapter II pointed out that alternate time-frequency representations allow computing time varying spectral energy features. In fact, Chapter III presented examples that may indicate the Wigner distribution produces better time varying spectral energy information than the spectrogram.

The remaining unanswered problem is that, although none of the reviewed time-frequency representations seem capable of replacing the spectrogram, does a time-frequency representation exist that can replace the spectrogram? To answer this

question, techniques that qualitatively and quantitatively compare different time-frequency representations must be developed. One comparative technique, Cohen's generalized time-frequency representation given by

$$P_x(t, f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(\theta, \tau) x\left(u + \frac{\tau}{2}\right) x^*\left(u - \frac{\tau}{2}\right) e^{-i2\pi(\theta t + \tau f - \theta u)} du d\tau d\theta, \quad (\text{IV.1})$$

represents different time-frequency representations using different kernel functions $\phi(\theta, \tau)$. By defining different continuous $\phi(\theta, \tau)$ functions, different time-frequency representations can be written in terms of $P_x(t, f)$. As an example, $\phi(\theta, \tau) = 1$ produces the Wigner distribution, while $\phi(\theta, \tau) = e^{-i\theta\tau^2}$ produces the Rihaczek distribution. Thus (IV.1) is capable of generalizing various time varying spectral energy techniques. Since Cohen's method generalizes time-frequency representations, analyzing kernel functions $\phi(\theta, \tau)$, allows comparing different time-frequency representations. However, although Cohen's generalized time-frequency representation allows proving time-frequency properties, little direct comparisons can be made by studying the different $\phi(\theta, \tau)$'s. One useful result that Cohen's method produces is that the Wigner distribution generates other time-frequency representations. This conclusion arises because $\phi(\theta, \tau)$ filters the $x(u + \tau/2)x^*(u - \tau/2)$ term, while $\phi(\theta, \tau) = 1$ implies a lack of filtering. This chapter introduces a new technique, the correlation approach for time-frequency representations, that produces a method for comparing different time-frequency representations. The time-frequency correlation approach produces new insights into different time-frequency representations.

Correlation Approach Background

Motivation

During the initial investigation of time-frequency representations, different methods for improving Wigner distribution performance were investigated. This investigation led

to the study of the $x(t + \tau/2)x^*(t - \tau/2)$ term found in the Wigner distribution,

$$W_x(t, f) = \int_{-\infty}^{\infty} x(t + \tau/2)x^*(t - \tau/2)e^{-i2\pi f\tau} d\tau. \quad (\text{IV.2})$$

The $x(t + \tau/2)x^*(t - \tau/2)$ term, called the *Wigner kernel*, is interesting because the remaining parts of (IV.2) are a Fourier transformation with respect to τ . Defining,

$$A_x(t, \tau) = x(t + \tau/2)x^*(t - \tau/2), \quad (\text{IV.3})$$

the Wigner distribution can now be expressed as

$$W_x = \int_{-\infty}^{\infty} A_x(t, \tau)e^{-i2\pi f\tau} d\tau. \quad (\text{IV.4})$$

Equation (IV.4) is the Fourier transformation of $A_x(t, \tau)$ with respect to τ .

Similarly, if other time-frequency representations formulate into (IV.4), then comparing different autocorrelation functions compares different time-frequency representations. Comparing time-frequency representations with autocorrelation functions is the basis of the correlation approach to time-frequency representations. The correlation approach to time-frequency representations has two advantages over Cohen's generalized time-frequency representation. First, engineers are more accustomed to autocorrelation functions as opposed to the $\phi(\theta, \tau)$ term in (IV.1), so the correlation approach provides better interpretations. Second, this chapter shows that the correlation approach allows stronger comparisons between different time-frequency representations and the spectrogram. In fact, the correlation approach provides an excellent forum for comparing the Wigner distribution to the spectrogram. However, before delving into the correlation approach, a short review is given on random process and how these random processes relate to spectral energies.

A Brief Random Process Theory Review

Throughout this thesis, the $x(t)$ term represents a signal. Representing a signal as $x(t)$ implies the signal is deterministic because the signal can be written down. However, any real world signal processing system introduces random noise into these otherwise deterministic signals. Thus, any real world signal, $x(t)$, is a deterministic signal plus random noise. As such, a signal $x(t)$ is just one realization from an infinite number of sample functions defined by a random process. Figure IV.1 illustrates these and more random process concepts. All the functions $x_i(t)$ given in Figure IV.1 comprise the ensemble of functions that describe a random process. The random process contains all of the possible signal values using a probability density function (pdf) that depends on time t . If t is set to a constant t_1 , then $x_i(t_1)$ is a random variable that depends on variable i whose values distribute according to the random process pdf. If i equals a constant k , then $x_k(t)$ is a sample function dependent on t . A sample function corresponds directly to a signal. Notice that the best scenario for signal knowledge in random systems is knowing a signal's probabilities at a specific time. In real world systems, complete statistical knowledge, such as the random process pdf, is not always available. Thus, the next best information consists of average signal values for a particular time and for how far signal values wander from the mean. These two quantities are called the first and second order moments of a random process. Recall that first order moments were computed on both the Wigner distribution and the spectrogram. In statistical terms, the first order moment is called the mean value and is defined by

$$\mu_x(t) = E[x_i(t)] = \int_{-\infty}^{\infty} \xi f_{x(t)}(\xi, t) d\xi, \quad (\text{IV.5})$$

where $f_x(\xi, t)$ is the random process pdf for a particular time and $E[\cdot]$ represents the expectation operator. Notice the similarity to the Wigner distribution's first order

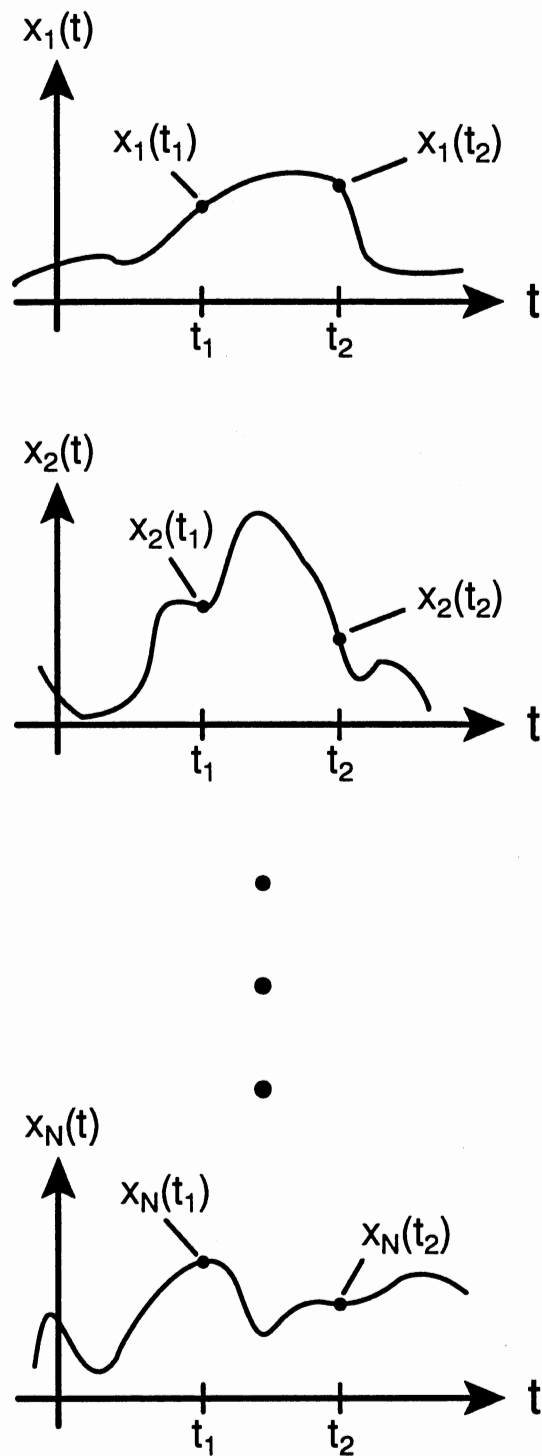


Figure IV.1 A Random Process Example

moment that was given in (III.35) as

$$m_1(t) = \int_{-\infty}^{\infty} \xi W_x(t, \xi) d\xi. \quad (\text{IV.6})$$

Thus, the pdf can be related directly to the Wigner distribution. This correspondence makes sense because $W_x(t, f)$ states how spectral energy is distributed while $f_x(\xi, t)$ indicates how probabilities distribute. The only problem with (IV.5) is that $f_x(\xi, t)$ rarely is known. Thus, $E[x_i(t)]$ is estimated using the ensemble average

$$\mu_x(t) = E[x_i(t)] = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t), \quad (\text{IV.7})$$

where $x_i(t)$ is a random variable. Note that (IV.7) is an unbiased estimator with an error variance of σ_x^2/N where σ_x^2 is the random variable variance [Ben86]. However, $\mu_x(t)$ is usually approximated by

$$\hat{\mu}_x(t) \equiv E[x_i(t)] \equiv \frac{1}{N} \sum_{i=1}^N x_i(t), \quad (\text{IV.8})$$

where N is assumed to finite to reduce computation time or storage requirements. Estimating the time dependent mean value is illustrated in Figure IV.1 by setting $t = t_1$ and averaging the points $x_i(t_1)$ for $i \in [1, \dots, N]$.

The second order moment, called the autocorrelation function, is defined by

$$R_x(t_1, t_2) = E[x_i(t_1)x_i^*(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \xi_1 \xi_2^* f_{x(t_1)x(t_2)}(\xi_1, \xi_2, t_1, t_2) d\xi_1 d\xi_2, \quad (\text{IV.9})$$

where $f_{x(t_1)x(t_2)}(\xi_1, \xi_2, t_1, t_2)$ is the random process pdf for times t_1 and t_2 . Similar to estimating a mean value, the ensemble average,

$$R_x(\hat{t}_1, \hat{t}_2) \equiv E[x_i(t_1)x_i^*(t_2)] \equiv \frac{1}{N} \sum_{i=1}^N x_i(t_1)x_i^*(t_2), \quad (\text{IV.10})$$

estimates the autocorrelation function. As with the ensemble mean average, this ensemble average is unbiased and produces better estimates as N gets large [Ben86]. Figure IV.1 illustrates estimating $R_x(t_1, t_2)$ by averaging the product of the $x_i(t_1)x_i(t_2)$ terms over all $i \in [1, \dots, N]$.

Thus, assuming random signals complicates signal processing. Not only are signal values not known in advance, but the ideal random information (the random process pdf) is almost never known. Thus, moments provide the only information available for signal processing. Although ensemble average moments do not seem capable of yielding much signal information, later in this section a spectral energy measure will be related to an autocorrelation function; and therefore, the ensemble average moments are shown to be useful.

The definitions of $\mu_x(t)$ and $R_x(t_1, t_2)$, equations (IV.6) and (IV.9), are both functions of time. Random processes that depend on time are called nonstationary. Nonstationary random processes are the random signal equivalent to time varying deterministic signals. Although the terms nonstationary and time varying are interchanged frequently in literature, this thesis uses the terms as defined here. Thus, nonstationary signals are random and vary in time while time varying signals are deterministic and vary with time.

The random process equivalent of deterministic signals is now discussed. In random signal theory, a random process that does not depend on time is called *strict-sense stationary*. Thus, a strict-sense stationary random variable $x_i(t)$ contains the identical statistical properties as $x_i(t - \xi)$ for all $\xi \in \mathcal{R}$. Because strict-sense stationary random processes rarely exist in the real world, a more restrictive random classification called *wide-sense stationary* is often used. A *wide-sense stationary* random process contains a constant mean value and an autocorrelation function that depends only on a time difference. Thus, the only difference between a wide-sense stationary random process and a strict-sense stationary random process is that the wide-sense stationary random process's

autocorrelation function depends on time while the strict-sense stationary random process's autocorrelation function does not. Typically, wide-sense stationary random processes are called stationary random processes. By definition, a wide-sense stationary random process has the following properties

$$\mu_x(t) = \mu_x, \quad \forall t \in \mathfrak{R} \quad (\text{IV.11})$$

and
$$R_x(t_1, t_2) = R_x(t, t - \tau) = R_x(\tau), \quad \forall t, \tau \in \mathfrak{R}. \quad (\text{IV.12})$$

Note that the independence of t in (IV.12) relates to the fact that the time chosen to compute the autocorrelation function is irrelevant. These two properties can be visualized by looking at Figure IV.1. Notice that (IV.11) and Figure IV.1 imply the mean can be estimated for any t . Similarly, sliding t_1 and t_2 simultaneously to any position on the t -axis of Figure IV.1 estimates the autocorrelation function.

Although not evident, assuming stationary random processes allows analyzing a signal's spectral energy. The connection between stationary random processes and spectral energies is discussed in this section after discussing an additional random process restriction. Regardless of whether a random process is stationary or nonstationary, ensemble averaged mean and the autocorrelation function estimates are desired. However, in a real world signal processing implementation, ensemble averages require many different copies of the same signal. As an example, in an organ melody identifying system, computing ensemble averages over a particular melody requires playing and storing the melody many times. Unfortunately, obtaining several signals is not possible in many systems either due to the nature of the signal or due to hardware constraints. Thus, many signal processing systems provide only one signal or sample function for analysis.

Suppose only one sample function, $x_k(t)$, is available from a random process $\{x_i(t), i \in Z^+\}$. Then the ensemble averages, for the random process, are given by

$$\mu_x \hat{(t)} \equiv \frac{1}{N} \sum_{i=1}^N x_i(t) = x_k(t) \quad (\text{IV.13})$$

and

$$R_x(\hat{t}_1, t_2) \equiv \frac{1}{N} \sum_{i=1}^N x_i(t_1) x_i^*(t_2) = x_k(t_1) x_k^*(t_2). \quad (\text{IV.14})$$

Note that although the estimates of $\mu_x(t)$ and $R_x(t_1, t_2)$ are unbiased, the error variances are high since the averages contain only one term. Thus, estimates computed by (IV.13) and (IV.14) are very unreliable. Producing some method for estimating the mean and autocorrelation function requires thinking of the sample function $x_k(t)$ as a deterministic signal. Thus, moments are computed using a time average defined by

$$\langle f(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(\xi) d\xi, \quad (\text{IV.15})$$

where $f(t)$ is any continuous function and $\langle \cdot \rangle$ denotes the time average. Thus, a deterministic mean or a time average mean is defined as

$$\langle \mu_x(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \mu_x(\xi) d\xi. \quad (\text{IV.16})$$

Substituting the ensemble average of $\mu_x(t)$, given in (IV.13), into (IV.16) produces

$$\langle \mu_x(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_k(\xi) d\xi, \quad (\text{IV.17})$$

for sample function number k .

Because the nonstationary autocorrelation function $R_x(t_1, t_2)$ is a function of two variables, computing a deterministic nonstationary autocorrelation function using (IV.15) is not possible. However, if the random process is assumed to be wide-sense stationary, then $R_x(t_1, t_2) = R_x(\tau)$ where $\tau = t_1 - t_2$ and the ensemble average for one sample function $x_k(t)$ produces

$$R_x(t, \hat{t} - \tau) \equiv \frac{1}{N} \sum_{i=1}^N x_i(t) x_i^*(t - \tau) = x_k(t) x_k^*(t - \tau), \quad (\text{IV.18})$$

for any $t \in \mathfrak{R}$. Thus, the deterministic autocorrelation function or the time averaged autocorrelation function is defined as

$$\langle R_x(t, t - \tau) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T R_x(\xi, \xi - \tau) d\xi. \quad (\text{IV.19})$$

Substituting the ensemble average of $R_x(t, t - \tau)$ from (IV.18) into the time average in (IV.19) yields

$$\langle R_x(\tau) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x_k(\xi) x_k^*(\xi - \tau) d\xi. \quad (\text{IV.20})$$

Time average estimates are important in signal processing for several reasons. First, sometimes time average values equal ensemble average values. When time averages equal ensemble averages with probability one, then the random process is called *ergodic*. Note that time averages are not a function of time so ergodic random processes are automatically stationary. However, many stationary random processes are not ergodic. Thus, time averages provide a convenient method of computing averages although time averages may not relate to ensemble averages.

The second reason time averages are so important is that for non-ergodic random processes, time averages sometimes provide more useful mean and autocorrelation function values when using only one sample function. For truly nonstationary random processes, neither one sample function ensemble averages nor time averages provide very useful results. However, for random processes that do not vary significantly over time, time averages can produce useful mean and autocorrelation values.

All the discussion concerning random processes and random process classifications would not belong in this thesis unless random processes related somehow to spectral energies. To see the relationship between random processes and spectral energies, let $x_k(t)$ represent a random signal or a sample function for some k . If $x_k(t)$ is deterministic, then

$$|X(f)|^2 = \left| \int_{-\infty}^{\infty} x_k(t) e^{-i2\pi f t} dt \right|^2 \quad (\text{IV.21})$$

would compute the spectral energy. However, for $x_k(t)$ random, (IV.21) is not integrable since

$$\lim_{r \rightarrow \infty} \int_{-r}^r x_k(t) e^{-i2\pi f t} dt \quad (\text{IV.22})$$

may not exist because the limit of $x_k(t)$ as $t \rightarrow \infty$ may not equal zero. Thus, formulating a random signal spectral energy in terms of the deterministic Fourier transformation is not plausible. However, if the limits of integration in (IV.21) are finite, the spectral energy can be evaluated if $x_k(t)$ is bounded. Recall that limiting signal values is similar to procedures discussed in Chapter II for the spectrogram. The spectrogram was defined as

$$S_x(t, f) = \left| \int_{-\infty}^{\infty} x(\xi) h(t - \xi) e^{-i2\pi f \xi} d\xi \right|^2, \quad (\text{IV.23})$$

where $h(t)$ is a window function. Relating the spectrogram to a random spectral energy calculation is done by setting

$$h(t) = \begin{cases} 0, & t < -T \\ 1, & -T \leq t \leq T \\ 0, & t > T \end{cases} \quad (\text{IV.24})$$

Thus,

$$|X(f)|^2 = S_x(t, f) = \left| \int_{-T}^T x_k(\xi) e^{-i2\pi f \xi} d\xi \right|^2, \quad (\text{IV.25})$$

where $x_k(\xi)$ has been substituted for the deterministic $x(t)$ of the spectrogram and has been assumed to be real valued. Splitting (IV.25) into two separate integrals yields

$$\begin{aligned}
|X(f)|^2 &= \left(\int_{-T}^T x_k(\xi) e^{-i2\pi f \xi} d\xi \right) \left(\int_{-T}^T x_k(\alpha) e^{-i2\pi f \alpha} d\alpha \right)^* \\
&= \int_{-T}^T \int_{-T}^T x_k(\xi) x_k^*(\alpha) e^{-i2\pi f(\xi - \alpha)} d\alpha d\xi.
\end{aligned} \tag{IV.26}$$

Using the transformations $\tau = \xi - \alpha$ and $t = \xi$ and the fact that the Jacobian $\left| \frac{\partial(\xi, \alpha)}{\partial(\tau, t)} \right| = 1$,

(IV.26) manipulates into

$$|X(f)|^2 = \int_{-T}^T \int_{\frac{\tau-T}{2}}^{\frac{\tau+T}{2}} x_k(t) x_k^*(t - \tau) e^{-i2\pi f \tau} dt d\tau. \tag{IV.27}$$

Because $x_k(t)$ is a random variable, nothing can reduce (IV.27) to a point where spectral energy estimates are computable. Thus, the $x_k(t)x_k^*(t - \tau)$ term must be replaced with something else. The first approach assumes the random process $\{x_i(t)\}$ is ergodic so that statistical properties are computable on one sample function. Regardless which of the two assumptions is made, (IV.27) can be rewritten as

$$|X(f)|^2 = \int_{-T}^T \left\{ \int_{\frac{\tau-T}{2}}^{\frac{\tau+T}{2}} x(t) x^*(t - \tau) dt \right\} e^{-i2\pi f \tau} d\tau, \tag{IV.28}$$

where the subscripts on $x_k(t)$ have been omitted to indicate a deterministic signal or any ergodic random process sample function. The quantity inside the parenthesis of (IV.28) is similar to the time average autocorrelation function defined in (IV.20). However, two things are missing from the

$$\int_{\frac{\tau-T}{2}}^{\frac{\tau+T}{2}} x(t) x(t - \tau) dt \tag{IV.29}$$

term that allows replacing (IV.29) with the time average autocorrelation function. First, (IV.29) requires $T \rightarrow \infty$ as a limit. Second, (IV.29) needs a factor of $1/T$. Applying the limit operation and including the $1/T$ factor to (IV.29) yields

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{\frac{\tau-T}{2}}^{\frac{\tau+T}{2}} x(t)x^*(t-\tau)d\tau = \langle R_x(\tau) \rangle. \quad (\text{IV.30})$$

Thus,

$$\int_{\frac{\tau-T}{2}}^{\frac{\tau+T}{2}} x(t)x^*(t-\tau)dt = T \langle R_x(\tau) \rangle, \quad (\text{IV.31})$$

if the limit operation is assumed. Substituting (IV.31) into (IV.28) and taking the required limit produces

$$|X(f)|^2 = \lim_{T \rightarrow \infty} \int_{-T}^T \{T \langle R_x(\tau) \rangle\} e^{-i2\pi f\tau} d\tau. \quad (\text{IV.32})$$

Typically, (IV.32) is rewritten as

$$\frac{1}{T} |X(f)|^2 = \lim_{T \rightarrow \infty} \int_{-T}^T \langle R_x(\tau) \rangle e^{-i2\pi f\tau} d\tau, \quad (\text{IV.33})$$

where $\frac{1}{T} |X(f)|^2$ is called the *periodogram*. Including the limit in the integration symbol,

the periodogram becomes

$$\frac{1}{T} |X(f)|^2 = \int_{-\infty}^{\infty} \langle R_x(\tau) \rangle e^{-i2\pi f\tau} d\tau \quad (\text{IV.34})$$

or just

$$\frac{1}{T} |X(f)|^2 = F[\langle R_x(\tau) \rangle], \quad (\text{IV.35})$$

where $F[\cdot]$ is the Fourier transformation. Notice that $|X(f)|^2 = S_x(t, f)$ implies that a rectangular windowed spectrogram is proportional to the Fourier transformation of the nonstationary autocorrelation function. The periodogram, (IV.34), is important because it produces spectral energy estimates given a time average autocorrelation function.

The second approach for evaluating (IV.27) requires assuming a stationary random process. Because an ergodic assumption restricts a random process more than a stationary assumption, the spectral energy produced using the stationary assumption should contain more accurate spectral energy estimates. Assuming a stationary random process implies that the expected value of the $x(t)x^*(t - \tau)$ term of (IV.27) equals $R_x(\tau)$. Thus, applying the expected value operator to (IV.26) obtains a usable form of (IV.27). Computing the expected value of (IV.27) produces

$$E[|X(f)|^2] = \int_{-T}^T \int_{\frac{\tau-T}{2}}^{\frac{\tau+T}{2}} R_x(\tau) e^{-i2\pi f\tau} dt d\tau. \quad (\text{IV.36})$$

Equation (IV.36) can be rewritten as

$$E[|X(f)|^2] = \int_{-T}^T \left\{ \int_{\frac{\tau-T}{2}}^{\frac{\tau+T}{2}} dt \right\} R_x(\tau) e^{-i2\pi f\tau} d\tau, \quad (\text{IV.37})$$

so that

$$\begin{aligned} E[|X(f)|^2] &= \int_{-T}^T \left(\frac{\tau+T}{2} - \frac{\tau-T}{2} \right) R_x(\tau) e^{-i2\pi f\tau} d\tau \\ &= \int_{-T}^T T R_x(\tau) e^{-i2\pi f\tau} d\tau. \end{aligned} \quad (\text{IV.38})$$

Moving the factor of T to the left side of (IV.38) and applying a limit operation produces

$$E\left[\frac{1}{T}|X(f)|^2\right] = \lim_{T \rightarrow \infty} \int_{-T}^T R_x(\tau) e^{-i2\pi f\tau} d\tau. \quad (\text{IV.39})$$

Notice that the left side of (IV.39) is the expected value of the periodogram, and the right side is the Fourier transformation of the stationary autocorrelation function. Thus, (IV.39) states that

$$\mathbb{E}\left[\frac{1}{T}|X(f)|^2\right] = F[R_x(\tau)] \quad (\text{IV.40})$$

where $\mathbb{E}\left[\frac{1}{T}|X(f)|^2\right]$ is called the *power spectral density* (note that the term density is not dropped here to maintain the quantity's true title). Therefore, the power spectral density is the expected value of the periodogram. The power spectral density is denoted by

$$S_x(f) = \mathbb{E}\left[\frac{1}{T}|X(f)|^2\right] \quad (\text{IV.41})$$

and is related to the stationary autocorrelation function by the Wiener-Knitchine relations

$$\begin{aligned} S_x(f) &= \int_{-\infty}^{\infty} R_x(\tau) e^{-i2\pi f\tau} d\tau \\ R_x(\tau) &= \int_{-\infty}^{\infty} S_x(f) e^{i2\pi f\tau} df. \end{aligned} \quad (\text{IV.42})$$

This section shows that if a signal is random, then spectral energy quantities can still be estimated. Estimating random signal spectral energies requires assuming the underlying random process is stationary or ergodic. Assuming a stationary random process allows computing the power spectral density as the Fourier transformation of the stationary autocorrelation function. If an ergodic random process is assumed, then the Fourier transformation of the time average autocorrelation function produces the periodogram. The point is that under different circumstances, spectral energy estimates arise from computing the Fourier transformation of some autocorrelation function. The next section introduces the idea of analyzing time-frequency representations by analyzing time-frequency autocorrelation functions.

Correlation Approach to Time-Frequency Representations

As stated at the beginning of this chapter, the correlation approach to time-frequency representations began by rewriting the Wigner distribution as

$$W_x(t, f) = \int_{-\infty}^{\infty} A_x(t, \tau) e^{-i2\pi f\tau} d\tau,$$

where $A_x(t, \tau) = x(t + \tau/2)x^*(t - \tau/2).$ (IV.43)

However, another motivating factor was the relationship between the periodogram to the time average autocorrelation function and the power spectral density to the ensemble average autocorrelation function. Here are three different spectral energy functions generated using a Fourier transformation of an autocorrelation function. The obvious method of analyzing each of these three techniques is to analyze their respective autocorrelation functions. However, because the autocorrelation functions are defined differently, comparing the autocorrelation functions is not easy. The difficulty can be avoided by using a nonstationary autocorrelation function which allows analyzing these three autocorrelation functions and many more. This is one of the main contributions of this research.

The correlation approach to time-frequency representations begins by defining a nonstationary autocorrelation function that generalizes all autocorrelation functions. The time-frequency autocorrelation function is defined by

$$A_x(t_1, t_2) \stackrel{\Delta}{=} E_A[x(t_1)x^*(t_2)], \quad (\text{IV.44})$$

where $E_A[\cdot]$ denotes the *time-frequency expected value* and $A_x(t_1, t_2)$ represents the *time-frequency autocorrelation function*. The time-frequency expected value equals the normal expected value except no regard is given towards random process assumptions. Thus, the time-frequency expected value can be estimated using time averages, ensemble averages, or any other averages. More on this will be presented later. Allowing different

expected value formulae may sound unusual but if a random process is actually nonstationary, then the normal ergodic and ensemble assumptions fall apart. Thus, no harm occurs in defining anything as an autocorrelation function for nonstationary signals. Defining an expectation operator loosely produces many different autocorrelation functions.

The time-frequency autocorrelation function is identical to a nonstationary autocorrelation function

$$R_x(t_1, t_2) = E[x(t_1)x^*(t_2)], \quad (\text{IV.45})$$

except for the different expectation operators. Since the time-frequency expected value is less restricted compared to the normal expected value,

$$R_x(t_1, t_2) \subseteq A_x(t_1, t_2) \quad (\text{IV.46})$$

for a particular random process $\{x_i(t)\}$. Before describing time-frequency representations in terms of the time-frequency autocorrelation function, some time-frequency autocorrelation function properties are discussed next.

The first thing to notice about the time-frequency autocorrelation function is its dependence on t_1 and t_2 . Depending on two time variables does not fit well with normal autocorrelation functions that are a function of a time t and a time difference τ . Recall that the periodogram and the power spectral density used $R(t, \tau)$ in their respective derivations. In addition, the Wigner distribution was rewritten in terms of an autocorrelation function that depended on t and τ in (IV.43). Thus, converting the time-frequency autocorrelation function into a function depending on t and τ is desirable. Converting variables t_1 and t_2 into t and τ requires a simple linear transformation. Thus, letting

$$\begin{aligned} t &= t_1 \\ \tau &= t_1 - t_2 \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} t_1 &= t \\ t_2 &= t - \tau \end{aligned} \quad (\text{IV.47})$$

allows writing the time-frequency autocorrelation function as

$$A_x(t_1, t_2) = A_x(t, t - \tau) = A_x(t, \tau). \quad (\text{IV.48})$$

Note that a slight notational change occurs in (IV.48). The $A_x(t_1, t_2) = A_x(t, t - \tau)$ term allows writing down the time-frequency expected value as $E[x(t_1)x^*(t_2)] = E[x(t)x^*(t - \tau)]$. However, placing the dependant time-frequency autocorrelation function variables directly into the time-frequency expected value does not work for $A_x(t, \tau)$ since $A_x(t, \tau) \neq E[x(t)x^*(\tau)]$. The $A_x(t, \tau)$ term requires knowledge of the transformation to write down the time-frequency expected value. Thus from now on, if the time-frequency autocorrelation function is given without a transformation then the dependent variables substitute for t_1 and t_2 directly. If a transformation accompanies the time-frequency autocorrelation function, then the dependent variables must be transformed before substituting into the time-frequency expected value.

In addition to possibly providing a transformation of dependent variables, the time-frequency expected value must also be stated. As an example, the stationary autocorrelation function can be written in terms of the time-frequency expected value by using the transformation in (IV.47) and by letting

$$A_x(t, \tau) = E_A[x(t)x^*(t - \tau)] = E[x(t)x^*(t - \tau)]. \quad (\text{IV.49})$$

Thus, the ensemble average autocorrelation function defines the time-frequency expected value. The time-frequency expected value could have been defined as the time average autocorrelation function. The point is that the time-frequency expected value contains different properties depending on the method used.

The time-frequency autocorrelation function generalizes many of the autocorrelation functions that define spectral energies. As such, a generalized spectral energy can be defined directly from the time-frequency autocorrelation function. The *time-frequency spectral energy* is defined as

$$TF_x(t, f) = \int_{-\infty}^{\Delta} A_x(t, \tau) e^{-i2\pi f\tau} d\tau, \quad (\text{IV.50})$$

where $A_x(t, \tau)$ denotes a finite energy continuous function and $TF_x(t, f)$ represents a time varying spectral energy. The reader should note that once again the term density has once again been dropped for consistency. Thus, the time-frequency spectral energy is actually a time-frequency spectral energy density. The time-frequency spectral energy is not unlike the other time varying spectral energies already discussed in this thesis because all are based upon computing the Fourier transformation of an autocorrelation function. However, because the time-frequency autocorrelation function defines the time-frequency spectral energy, the time-frequency spectral energy is capable of describing many different spectral energy techniques. Different spectral energies are placed into the context of the time-frequency spectral energy by defining the proper time-frequency autocorrelation function. As an example, the periodogram and the power spectral density will be written in terms of the time-frequency spectral energy.

Recall that the periodogram is defined as

$$\frac{1}{T} |X(f)|^2 = \int_{-\infty}^{\infty} \langle R_x(\tau) \rangle e^{-i2\pi f\tau} d\tau, \quad (\text{IV.51})$$

where $\langle R_x(\tau) \rangle$ represents the time-frequency autocorrelation function, and

$$\begin{aligned} \langle R_x(\tau) \rangle &= E[x(t)x^*(t - \tau)] \\ &= \int_{-\infty}^{\infty} x(\xi)x^*(\xi - \tau)d\xi. \end{aligned} \quad (\text{IV.52})$$

The time-frequency expected value is defined as

$$E_A[x(t)x^*(t - \tau)] = \int_{-\infty}^{\infty} x(\xi)x^*(\xi - \tau)d\xi. \quad (\text{IV.53})$$

Thus, the transformation (IV.47) and the time average time-frequency expected value in (IV.53) defines the time-frequency autocorrelation function, $A_x(t, \tau)$. Therefore, the time-frequency spectral energy becomes

$$\begin{aligned}
 TF_x(t, f) &= \int_{-\infty}^{\infty} A_x(t, \tau) e^{-i2\pi f\tau} d\tau \\
 &= \frac{1}{T} |X(f)|^2,
 \end{aligned} \tag{IV.54}$$

where transformation (IV.47) and the time average expected value from (IV.52) define $A_x(t, \tau)$. Thus, the time-frequency spectral energy easily describes the periodogram.

Next, the power spectral density is described using the time-frequency spectral energy.

Recall that the power spectral density is defined by

$$S_x(f) = \int_{-\infty}^{\infty} R_x(\tau) e^{-i2\pi f\tau} d\tau, \tag{IV.55}$$

where $R_x(\tau)$ is the stationary autocorrelation function defined using ensemble averages.

Because,

$$\begin{aligned}
 R_x(\tau) &= E[x(t)x^*(t - \tau)] \\
 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i^*(t - \tau),
 \end{aligned} \tag{IV.56}$$

the time-frequency expected value is defined as

$$E_A[x(t)x^*(t - \tau)] = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i^*(t - \tau). \tag{IV.57}$$

Thus, transformation (IV.47) and the ensemble average time-frequency expected value given in (IV.56) define $A_x(t, \tau)$. Therefore, the time-frequency spectral energy becomes

$$\begin{aligned}
 TF_x(t, f) &= \int_{-\infty}^{\infty} A_x(t, \tau) e^{-i2\pi f\tau} d\tau \\
 &= S_x(f),
 \end{aligned} \tag{IV.58}$$

where $A_x(t, \tau)$ uses transformation (IV.47) and the ensemble average time-frequency

expected value in (IV.56). Thus, the time-frequency spectral energy describes the power spectral density.

Although showing that the time-frequency spectral energy generalizes both the periodogram and the power spectral density was not difficult, the examples do indicate the potential power of the time-frequency spectral energy. The time-frequency spectral energy is powerful because defining different time-frequency autocorrelation functions produced two different spectral energies that arise in signal processing. Next, by defining a different transformation, the Wigner distribution is written in terms of the time-frequency spectral energy.

At the beginning of this section, the Wigner distribution was written in terms of the Fourier transformation of a generic autocorrelation function. This generic autocorrelation function is replaced by the time-frequency autocorrelation function. Because the Wigner distribution is defined by

$$W_x(t, f) = \int_{-\infty}^{\infty} x(t + \tau/2) x^*(t - \tau/2) e^{-i2\pi f\tau} d\tau, \quad (\text{IV.59})$$

it follows that

$$A_x(t, \tau) = x(t + \tau/2) x^*(t - \tau/2). \quad (\text{IV.59a})$$

As always, two items must be defined for the time-frequency autocorrelation function.

First, since $A_x(t, \tau)$ depends on t and τ , a transformation for t_1 and t_2 is needed. Letting

$$\begin{aligned} t &= \frac{t_1 + t_2}{2} & \Leftrightarrow & & t_1 &= t + \frac{\tau}{2} \\ \tau &= t_1 - t_2 & & & t_2 &= t - \frac{\tau}{2} \end{aligned} \quad (\text{IV.60})$$

defines the time-frequency autocorrelation function transformation and allows writing

$$\begin{aligned} A_x(t_1, t_2) &= E_A[x(t_1) x^*(t_2)] \\ &= E_A[x(t + \tau/2) x^*(t - \tau/2)] \\ &= A_x(t, \tau). \end{aligned} \quad (\text{IV.61})$$

Next, because it is desired that $A_x(t, \tau) = x(t + \tau/2)x^*(t - \tau/2)$, the time-frequency expected value is defined so that

$$\begin{aligned} E_A[x(t_1)x^*(t_2)] &= x(t_1)x^*(t_2) \\ &= x(t + \tau/2)x^*(t - \tau/2). \end{aligned} \quad (\text{IV.62})$$

Defining the time-frequency expected value as in (IV.62) is equivalent to an ensemble average over one sample function (recall (IV.18)). Thus, defining $A_x(t, \tau)$ using transformation (IV.60) and a one sample function ensemble average time-frequency expected value allows writing

$$\begin{aligned} TF_x(t, f) &= \int_{-\infty}^{\infty} A_x(t, \tau) e^{-i2\pi f\tau} d\tau \\ &= W_x(t, f). \end{aligned} \quad (\text{IV.63})$$

Therefore, the Wigner distribution is easily written in terms of the time-frequency spectral energy by means of the time-frequency autocorrelation function.

Now that the three different spectral energies (the periodogram, the power spectral density, and the Wigner distribution) have been written in terms of the time-frequency spectral energy, the three spectral energies can be compared. Since the only difference between the three spectral energies is differing time-frequency expected values and time-frequency autocorrelation functions, Table I helps compare these techniques. The main item provided by Table I is that three other time-frequency spectral energies exist within the context of transformations (IV.47) and (IV.60) when considering the time-frequency expected values already discussed. Before comparing the three time-frequency spectral energies, the missing time-frequency spectral energies designated as question marks in Table I are investigated.

TABLE I
A COMPARISON OF THREE TIME-FREQUENCY
SPECTRAL ENERGIES

	Transformation IV.47	Transformation IV.60
Time Average Time-Frequency Expected Value	Periodogram	?
Ensemble Average Time-Frequency Expected Value	Power Spectral Density	?
One Sample Function Ensemble Average Time-Frequency Expected Value	?	Wigner Distribution

The first missing time-frequency spectral energy defines the time-frequency autocorrelation function with transformation (IV.60) and the time average time-frequency expected value. Thus, the time-frequency autocorrelation function becomes

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t + \tau/2)x^*(t - \tau/2)] \\ &= \int_{-\infty}^{\infty} x(\xi + \tau/2)x^*(\xi - \tau/2)d\xi. \end{aligned} \quad (\text{IV.64})$$

Before inserting this time-frequency autocorrelation function into the time-frequency spectral energy equation (IV.50), (IV.64) is manipulated in an attempt to relate the time-frequency spectral energy to a well known spectral energy technique. Although such a relationship may not exist, the three spectral energies already examined did directly relate to known spectral energies.

The time-frequency autocorrelation function equation (IV.64) is modified by performing the transformation of $\xi = t + \tau/2$ inside the integrand. Because the transformation implies that $t = \xi - \tau/2$ and $\frac{\partial \xi}{\partial t} = 1$, the time-frequency autocorrelation function becomes

$$\begin{aligned} A_x(t, \tau) &= \int_{-\infty}^{\infty} x(\xi)x^*(\xi - \tau)d\xi \\ &= \langle R_x(\tau) \rangle. \end{aligned} \quad (\text{IV.65})$$

Therefore, the time-frequency autocorrelation function defined by transformation (IV.60) and the time average time-frequency expected value equals the time average autocorrelation function (IV.20). Therefore,

$$\begin{aligned} TF_x(t, f) &= \int_{-\infty}^{\infty} A_x(t, \tau)e^{-i2\pi f\tau}d\tau \\ &= \frac{1}{T}|X(f)|^2, \end{aligned} \quad (\text{IV.66})$$

or the time-frequency spectral energy, (IV.50), equals the periodogram.

The second missing time-frequency spectral energy defines the time-frequency autocorrelation function with transformation (IV.60) and an ensemble average time-frequency expected value. Thus, the time-frequency autocorrelation function becomes

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t + \tau/2)x^*(t - \tau/2)] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t + \tau/2)x_i^*(t - \tau/2). \end{aligned} \quad (\text{IV.67})$$

Again, $A_x(t, \tau)$ is manipulated to resemble an established spectral energy procedure.

Substituting ξ for $t + \tau/2$ modifies the time-frequency autocorrelation function. Because $t = \xi - \tau/2$ and $\frac{\partial \xi}{\partial t} = 1$, the time-frequency autocorrelation function becomes

$$\begin{aligned} A_x(\xi, \tau) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x(\xi)x^*(\xi - \tau) \\ &= E[x(\xi)x^*(\xi - \tau)] \\ &= R_x(\xi, \tau). \end{aligned} \quad (\text{IV.68})$$

Therefore, the time-frequency autocorrelation function defined by transformation (IV.60) and the ensemble average time-frequency expected value equals the ensemble average autocorrelation function. Additionally, if the random process is assumed stationary then

$$A_x(\xi, \tau) = R_x(\tau), \quad (\text{IV.68a})$$

which implies the time-frequency expected value equals the stationary autocorrelation function. The reader should note that (IV.68a) equals (IV.65) when a random process is ergodic, but they differ if the random process is only stationary. The time-frequency spectral energy is given by

$$\begin{aligned} TF_x(t, f) &= \int_{-\infty}^{\infty} A_x(t, \tau) e^{-i2\pi f\tau} d\tau \\ &= S_x(t, f), \end{aligned} \quad (\text{IV.69})$$

and equals

$$TF_x(t, f) = S_x(f) \quad (\text{IV.69a})$$

when the random process is stationary. Thus, the time-frequency spectral energy equals the power spectral density. As with the periodogram, the time-frequency spectral energy produces the power spectral density using either transformation (IV.47) or transformation (IV.60) and the ensemble average time-frequency expected value.

The third, and final, missing time-frequency spectral energy defines the time-frequency autocorrelation function with transformation (IV.47) and a one sample function ensemble average time-frequency expected value. Thus, the time-frequency autocorrelation function equals

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t)x^*(t - \tau)] \\ &= x(t)x^*(t - \tau). \end{aligned} \quad (\text{IV.70})$$

Unlike the two previous missing time-frequency spectral energies, the time-frequency spectral energy itself is manipulated and not the time-frequency autocorrelation function. Therefore, the time-frequency spectral energy becomes

$$\begin{aligned} TF_x(t, f) &= \int_{-\infty}^{\infty} A_x(t, \tau) e^{-i2\pi f\tau} d\tau \\ &= \int_{-\infty}^{\infty} x(t)x^*(t - \tau) e^{-i2\pi f\tau} d\tau \\ &= x(t) \int_{-\infty}^{\infty} x^*(t - \tau) e^{-i2\pi f\tau} d\tau. \end{aligned} \quad (\text{IV.71})$$

Letting $\xi = t - \tau$ implies $\tau = t - \xi$ and $\left| \frac{\partial \xi}{\partial \tau} \right| = 1$ so that

$$\begin{aligned}
TF_x(t, f) &= x(t) \int_{-\infty}^{\infty} x^*(\xi) e^{-i2\pi f(t-\xi)} d\xi \\
&= x(t) e^{-i2\pi ft} \int_{-\infty}^{\infty} x^*(\xi) e^{i2\pi f\xi} d\xi.
\end{aligned} \tag{IV.72}$$

Using conjugate function properties,

$$\begin{aligned}
TF_x(t, f) &= x(t) e^{-i2\pi ft} \left(\int_{-\infty}^{\infty} x(\xi) e^{-i2\pi f\xi} d\xi \right)^* \\
&= x(t) e^{-i2\pi ft} X^*(f).
\end{aligned} \tag{IV.73}$$

Recalling (II.73), notice that $TF_x(t, f)$ equals the Rihaczek distribution. Therefore, the time-frequency autocorrelation function defined with transformation (IV.47) and a one sample function ensemble average time-frequency expected value produces a time-frequency spectral energy equivalent to the Rihaczek distribution.

Now that the missing time-frequency spectral energies have been identified, Table I is updated and presented in Table II. Table II produces some interesting comparisons. First, the widely used periodogram and the power spectral density appear twice using two different transformations. Thus, the periodogram and the power spectral density seem robust to different time variable transformations. The same is not true, however, for single sample function ensemble average time-frequency expected value based spectral energies. By slightly modifying the time transformation, the Wigner distribution becomes the Rihaczek distribution. The link between the Wigner distribution and the Rihaczek distribution is surprising considering the Wigner distribution is real valued while the Rihaczek distribution is complex valued. However, recall that both the Wigner distribution and the Rihaczek distribution contain the instantaneous power and the power spectral density as marginal distributions, so the two are similar in some regard.

TABLE II
A COMPARISON OF SIX TIME-FREQUENCY
SPECTRAL ENERGIES

Transformation IV.47		Transformation IV.60
Time Average Time-Frequency Expected Value	Periodogram	Periodogram
Ensemble Average Time-Frequency Expected Value	Power Spectral Density	Power Spectral Density
One Sample Function Ensemble Average Time-Frequency Expected Value	Rihaczek Distribution	Wigner Distribution

Although much is already known about the properties contained in the four unique time-frequency spectral energies given in Table II, the following attempts to verify these properties using original time-frequency spectral energy properties. Recall that Cohen's Generalized time-frequency representation allowed studying different time varying spectral energies. In the same manor, the correlation approach allows investigating and comparing different time varying spectral energies. However, the correlation approach's time-frequency spectral energy and the time-frequency autocorrelation function allow easier development of different spectral energy methods and produce very useful analysis properties.

One of the more important properties answers the question of when time-frequency spectral energies are real valued. Knowing when a time-frequency spectral energy is real valued is important because complex valued time-frequency representations, such as the Rihaczek distribution, are difficult to interpret. The Properties IV.1 and IV.2 (given next) indicate when a time-frequency spectral energy is real valued. Before discussing time-frequency spectral energy properties, recall that the possibility of complex signals exists. Thus, all the time-frequency spectral energy properties must take into account both complex and real valued signals. In cases where the real signal property falls directly from the complex signal derivation, only the complex signal proof is presented.

Property IV.1: The time-frequency spectral energy is real valued if $A_x(t, \tau) = A_x^*(t, -\tau)$ and $A_x(t, \tau)$ is complex.

Proof: Using an integral property, the time-frequency spectral energy separates into

$$TF_x(t, f) = \int_0^{\infty} A_x(t, \tau) e^{-i2\pi f\tau} d\tau + \int_{-\infty}^0 A_x(t, \tau) e^{-i2\pi f\tau} d\tau.$$

The second integral is rewritten using the transformation $\xi = -\tau$ so that

$$\begin{aligned}\int_{-\infty}^0 A_x(t, \tau) e^{-i2\pi f\tau} d\tau &= \int_{\infty}^{-0} A_x(t, -\xi) e^{-i2\pi f(-\xi)} (-1) d\xi \\ &= \int_0^{\infty} A_x(t, -\xi) e^{i2\pi f\xi} d\xi.\end{aligned}$$

Since $A_x(t, -\xi) = A_x^*(t, \xi)$,

$$\begin{aligned}\int_{-\infty}^0 A_x(t, \tau) e^{-i2\pi f\tau} d\tau &= \int_0^{\infty} A_x^*(t, \xi) e^{i2\pi f\xi} d\xi \\ &= \left(\int_0^{\infty} A_x(t, \xi) e^{-i2\pi f\xi} d\xi \right)^*.\end{aligned}$$

$$\text{Hence, } TF_x(t, f) = \int_0^{\infty} A_x(t, \tau) e^{-i2\pi f\tau} d\tau + \left(\int_0^{\infty} A_x(t, \xi) e^{-i2\pi f\xi} d\xi \right)^*.$$

Because adding a complex number with its conjugate produces real values, $TF_x(t, f)$ is

real. †††

Property IV.2: The time-frequency spectral energy is real valued if $A_x(t, \tau) = A_x(t, -\tau)$ and

$A_x(t, \tau)$ is real.

Proof: Using Euler's Identity,

$$\begin{aligned}TF_x(t, f) &= \int_{-\infty}^{\infty} A_x(t, \tau) e^{-i2\pi f\tau} d\tau \\ &= \int_{-\infty}^{\infty} A_x(t, \tau) \cos(2\pi f\tau) d\tau - i \int_{-\infty}^{\infty} A_x(t, \tau) \sin(2\pi f\tau) d\tau.\end{aligned}$$

The integral of the complex part is rewritten using integral properties as

$$\int_{-\infty}^{\infty} A_x(t, \tau) \sin(2\pi f\tau) d\tau = \int_0^{\infty} A_x(t, \tau) \sin(2\pi f\tau) d\tau + \int_{-\infty}^0 A_x(t, \tau) \sin(2\pi f\tau) d\tau.$$

Applying the transformation $\xi = -\tau$ to the second integral yields

$$\begin{aligned} \int_{-\infty}^0 A_x(t, \tau) \sin(2\pi f \tau) d\tau &= \int_{\infty}^0 A_x(t, -\xi) \sin(2\pi f(-\xi)) (-1) d\xi \\ &= \int_0^{\infty} A_x(t, -\xi) (-1) \sin(2\pi f \xi) (-1) d\xi \\ &= - \int_0^{\infty} A_x(t, -\xi) \sin(2\pi f \xi) d\xi. \end{aligned}$$

Since $A_x(t, -\xi) = A_x(t, \xi)$,

$$\begin{aligned} \int_{-\infty}^{\infty} A_x(t, \tau) \sin(2\pi f \tau) d\tau &= \int_0^{\infty} A_x(t, \tau) \sin(2\pi f \tau) d\tau - \int_0^{\infty} A_x(t, \xi) \sin(2\pi f \xi) d\xi \\ &= 0. \end{aligned}$$

Hence,
$$TF_x(t, f) = \int_{-\infty}^{\infty} A_x(t, \tau) \cos(2\pi f \tau) d\tau + i(0).$$

Because $A_x(t, \tau)$ is real and no other imaginary term exists, $TF_x(t, f)$ is real valued. †††

Properties IV.1 and IV.2 indicate that if the time-frequency autocorrelation function is an even function with respect to τ , then the time-frequency spectral energy is real valued. This also occurs in conventional signal processing. Because the time-frequency autocorrelation function is defined by a transformation and a time-frequency expected value, the time-frequency spectral energy realness property depends on both the transformation and the time-frequency expected value. The following properties indicate which of the six time-frequency spectral energies studied so far produce real values.

Property IV.3: Using transformation (IV.47) and a time averaged time-frequency expected value, $A_x(t, \tau) = A_x^*(t, -\tau)$ for complex $x(t)$ and $A_x(t, \tau) = A_x(t, -\tau)$ for real $x(t)$.

Proof: Since $A_x(t, \tau)$ is defined with transformation (IV.47) and a time average time-frequency expected value,

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t)x^*(t - \tau)] \\ &= \langle x(t)x^*(t - \tau) \rangle \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(\alpha)x^*(\alpha - \tau) d\alpha. \end{aligned}$$

Similarly, $A_x(t, -\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(\alpha)x^*(\alpha + \tau) d\alpha.$

Substituting ξ for $\alpha + \tau$ implies that $\alpha = \xi - \tau$ and

$$A_x(t, -\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(\xi - \tau)x^*(\xi) d\xi.$$

$$\begin{aligned} \text{Thus, } A_x(t, -\tau) &= \left(\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x^*(\xi - \tau)x(\xi) d\xi \right)^* \\ &= A_x^*(t, \tau). \end{aligned}$$

Hence, $A_x(t, \tau) = A_x^*(t, -\tau)$ for complex $x(t)$. Additionally, $A_x(t, \tau) = A_x(t, -\tau)$ for real $x(t)$

because $A_x(t, \tau)$ is real valued for real $x(t)$. †††

Property IV.4: For real $x(t)$, $A_x(t, \tau)$ may not equal $A_x(t, -\tau)$ when the time-frequency autocorrelation function is defined with transformation (IV.47) and an ensemble averaged time-frequency expected value.

Proof: Since $A_x(t, \tau)$ is defined with transformation (IV.47) and an ensemble average time-frequency expected value,

$$\begin{aligned}
A_x(t, \tau) &= E_A[x(t)x^*(t - \tau)] \\
&= E[x(t)x(t - \tau)] \\
&= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i(t - \tau).
\end{aligned}$$

Similarly, $A_x(t, -\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i(t + \tau).$

Suppose $A_x(t, \tau) = A_x(t, -\tau).$

Let $x_i(t) = \begin{cases} 1, & t > 0 \\ -1, & t \leq 0 \end{cases}$

and $t = 0$ so that $x_i(t)x_i(t - \tau) = x_i(0)x_i(-\tau) = \begin{cases} 1, & \tau > 0 \\ -1, & \tau \leq 0 \end{cases}.$

If $\tau = 1$, then $A_x(0, \tau) = x(0)x(-1) = 1$ while $A_x(0, -\tau) = x(0)x(1) = -1$. Since $1 \neq -1$, the hypothesis that $A_x(t, \tau) = A_x(t, -\tau)$ is contradicted. †††

Property IV.5: For complex valued $x(t)$, $A_x(t, \tau)$ may not equal $A_x^*(t, -\tau)$ when the time-frequency autocorrelation function is defined with transformation (IV.47) and an ensemble average time-frequency expected value.

Proof: Since $A_x(t, \tau)$ is defined with transformation (IV.47) and an ensemble average time-frequency expected value, it follows that

$$\begin{aligned}
A_x(t, \tau) &= E_A[x(t)x^*(t - \tau)] \\
&= E[x(t)x^*(t - \tau)] \\
&= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i^*(t - \tau).
\end{aligned}$$

Similarly, $A_x(t, -\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i^*(t + \tau).$

Suppose $A_x(t, \tau) = A_x^*(t, -\tau).$

Let $x_i(t) = \begin{cases} 1, & t > 0 \\ -1, & t \leq 0 \end{cases}$

and $t = 0$ so that $x_i(t)x_i^*(t - \tau) = x_i(0)x_i^*(-\tau) = \begin{cases} 1, & \tau > 0 \\ -1, & \tau \leq 0 \end{cases}$.

If $\tau = 1$, then $A_x(0, \tau) = x(0)x^*(-1) = 1$ while $A_x(0, -\tau) = x(0)x^*(1) = -1$. Since $1 \neq -1$, the hypothesis that $A_x(t, \tau) = A_x^*(t, -\tau)$ is contradicted. †††

Property IV.6: If $x(t)$ is real valued and the random process is stationary, then

$A_x(t, \tau) = A_x(t, -\tau)$ for time-frequency autocorrelation functions using transformation (IV.47) and an ensemble averaged time-frequency expected value.

Proof: Since $A_x(t, \tau)$ is defined with transformation (IV.47) and an ensemble average time-frequency expected value,

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t)x^*(t - \tau)] \\ &= E[x(t)x(t - \tau)] \\ &= R_x(\tau). \end{aligned}$$

Similarly, $A_x(t, -\tau) = E[x(t)x(t + \tau)] = R_x(-\tau)$.

Let $\xi = t + \tau$ so that $t = \xi - \tau$ and

$$A_x(t, -\tau) = E[x(\xi - \tau)x(\xi)] = R_x(-\tau).$$

By assuming a stationary random process,

$$R_x(\tau) = E[x(\xi - \tau)x(\xi)] = E[x(t)x(t + \tau)] = R_x(-\tau)$$

implies $A_x(t, -\tau) = A_x(t, \tau)$.

Hence, the property is proved. †††

Property IV.7: If $x(t)$ is complex valued and the random process is stationary, then

$A_x(t, \tau) = A_x^*(t, -\tau)$ for time-frequency autocorrelation function using transformation (IV.47) and an ensemble averaged time-frequency expected value.

Proof: Since $A_x(t, \tau)$ is defined with transformation (IV.47) and an ensemble average time-frequency expected value,

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t)x^*(t - \tau)] \\ &= E[x(t)x^*(t - \tau)] \\ &= R_x(\tau). \end{aligned}$$

Similarly, $A_x(t, -\tau) = E[x(t)x^*(t + \tau)] = R_x(-\tau)$.

Let $\xi = t + \tau$ so that $t = \xi - \tau$ and

$$A_x(t, -\tau) = E[x(\xi - \tau)x^*(\xi)] = R_x^*(-\tau).$$

By assuming a stationary random process,

$$R_x^*(\tau) = E[x(\xi - \tau)x^*(\xi)] = E[x(t)x^*(t + \tau)] = R_x(-\tau)$$

implies $A_x(t, -\tau) = A_x^*(t, \tau)$.

Hence, the property is proved. †††

Property IV.8: If $x(t)$ is complex or real valued, then $A_x(t, \tau) = A_x^*(t, -\tau)$ when the time-frequency autocorrelation function is defined with transformation (IV.60) and a time average time-frequency expected value.

Proof: Since $A_x(t, \tau)$ uses transformation (IV.60) and a time average time-frequency expected value,

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t + \tau/2)x^*(t - \tau/2)] \\ &= \langle x(t + \tau/2)x^*(t - \tau/2) \rangle \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(\xi + \tau/2)x^*(\xi - \tau/2)d\xi. \end{aligned}$$

Similarly, $A_x(t, -\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(\xi - \tau/2)x^*(\xi + \tau/2)d\xi$

and

$$\begin{aligned}
 A_x^*(t, -\tau) &= \left(\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(\xi - \tau/2) x^*(\xi + \tau/2) d\xi \right)^* \\
 &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x^*(\xi - \tau/2) x(\xi + \tau/2) d\xi \\
 &= A_x(t, \tau).
 \end{aligned}$$

Hence, $A_x(t, \tau) = A_x^*(t, -\tau)$ for complex $x(t)$, and thus, $A_x(t, \tau) = A_x(t, -\tau)$ for real $x(t)$.

†††

Property IV.9: Regardless if $x(t)$ is complex or real valued, $A_x(t, \tau) = A_x^*(t, -\tau)$ when the time-frequency autocorrelation function is defined with transformation (IV.60) and an ensemble average time-frequency expected value.

Proof: Since $A_x(t, \tau)$ uses transformation (IV.60) and an ensemble average time-frequency expected value,

$$\begin{aligned}
 A_x(t, \tau) &= E_A[x(t + \tau/2)x^*(t - \tau/2)] \\
 &= E[x(t + \tau/2)x^*(t - \tau/2)] \\
 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x(t + \tau/2)x^*(t - \tau/2).
 \end{aligned}$$

Similarly, $A_x(t, -\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x(t - \tau/2)x^*(t + \tau/2)$.

Because $A_x^*(t, \tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x^*(t + \tau/2)x(t - \tau/2)$,

$A_x(t, \tau) = A_x^*(t, -\tau)$ for complex $x(t)$ and $A_x(t, \tau) = A_x(t, -\tau)$ for real $x(t)$. †††

TABLE III
A PROPERTY COMPARISON OF SIX TIME-FREQUENCY
SPECTRAL ENERGIES

	Transformation IV.47	Transformation IV.60
Time Average Time-Frequency Expected Value	$x(t) \in C \Rightarrow A_x(t, \tau) = A_x^*(t, -\tau)$ $x(t) \in \Re \Rightarrow A_x(t, \tau) = A_x(t, -\tau)$	$x(t) \in C \Rightarrow A_x(t, \tau) = A_x^*(t, -\tau)$ $x(t) \in \Re \Rightarrow A_x(t, \tau) = A_x(t, -\tau)$
Ensemble Average Time-Frequency Expected Value	$x(t) \in C \Rightarrow A_x(t, \tau) \neq A_x^*(t, -\tau)$ $x(t) \in \Re \Rightarrow A_x(t, \tau) \neq A_x(t, -\tau)$	$x(t) \in C \Rightarrow A_x(t, \tau) = A_x^*(t, -\tau)$ $x(t) \in \Re \Rightarrow A_x(t, \tau) = A_x(t, -\tau)$
Ensemble Average Time-Frequency Expected Value (stationary)	$x(t) \in C \Rightarrow A_x(t, \tau) = A_x^*(t, -\tau)$ $x(t) \in \Re \Rightarrow A_x(t, \tau) = A_x(t, -\tau)$	$x(t) \in C \Rightarrow A_x(t, \tau) = A_x^*(t, -\tau)$ $x(t) \in \Re \Rightarrow A_x(t, \tau) = A_x(t, -\tau)$
One Sample Func- tion Ensemble Average Time-Frequency Expected Value	$x(t) \in C \Rightarrow A_x(t, \tau) \neq A_x^*(t, -\tau)$ $x(t) \in \Re \Rightarrow A_x(t, \tau) \neq A_x(t, -\tau)$	$x(t) \in C \Rightarrow A_x(t, \tau) = A_x^*(t, -\tau)$ $x(t) \in \Re \Rightarrow A_x(t, \tau) = A_x(t, -\tau)$

Properties IV.3 through IV.9 provide a great deal of information concerning the time-frequency spectral energies found in Table II. To compare different time-frequency autocorrelation functions, Table III provides an review of these properties. First, Properties IV.3 and IV.8 indicate that time average time-frequency autocorrelation functions immediately produce real time-frequency spectral energies. These obviously produce real time-frequency spectral energies because time averages do not depend on time. Therefore, changing τ 's sign makes no difference in the time-frequency autocorrelation function. Also, note that real time-frequency spectral energies are expected since the magnitude operation exists in the periodogram.

The properties involving the ensemble average time-frequency expected value produce very different results than produced by the time average time-frequency expected value. Properties IV.4 and IV.5 indicate that using transformation (IV.47) and ensemble average time-frequency expected values does not necessarily produce real time-frequency spectral energies. This result is surprising since this time-frequency spectral energy represents the power spectral density. However, the power spectral density implicitly assumed the underlying random processes is stationary. When the ensemble average time-frequency expected value assumes stationary random processes along with using transformation (IV.47), Properties IV.6 and IV.7 state that the time-frequency spectral energy is real valued. This result confirms the fact that the power spectral density is real. One interesting result of Properties IV.4 through IV.7 is that ensemble average time-frequency expected values computed using transformation (IV.47) produce complex time-frequency spectral energies.

When the ensemble average time-frequency expected value's transformation is changed from (IV.47) to (IV.60), the time-frequency spectral energy's properties change. Notice that Property IV.9 indicates that the time-frequency spectral energy is always real when using transformation (IV.60) and an ensemble average time-frequency expected value. This result contradicts the previous conclusion which required a stationary ran-

dom process assumption. Although ensemble average time-frequency expected values and either transformation (IV.47) or (IV.60) produce the power spectral density, transformation (IV.60) might be preferable because real time-frequency spectral energies always result.

As for the one sample function ensemble average time-frequency expected value based time-frequency spectral energies, Properties IV.4 and IV.5 indicate that transformation (IV.47) produces complex values. Thus, the Rihaczek distribution is always complex valued. Alternately, using transformation (IV.60), the time-frequency spectral energy is real valued. Therefore, the Wigner distribution is always real. The different results obtained between transformation (IV.47) and transformation (IV.60) indicate again that when desiring real time-frequency spectral energies, transformation (IV.60) is the best choice.

Although having a real valued time-frequency spectral energy is quite desirable since it possibly can replace the spectrogram, complex valued time-frequency spectral energies may be useful for generating signal processing features. Chapter III indicated that the marginal distributions with respect to time might make useful features; thus, obtaining properties which indicate when a time-frequency spectral energy produces useful marginal distributions is quite desirable. The next set of properties address the issue of when a particular time-frequency spectral energy produces potentially useful time varying spectral energy information. The next two properties indicate when time-frequency spectral energies produce two possible features: the instantaneous power and the power spectral energy.

Property IV.10: A time-frequency spectral energy produces the instantaneous power as a marginal distribution with respect to frequency if $A_x(t, 0) = |x(t)|^2$.

Proof: Using the marginal distribution definition of instantaneous power,

$$|x(t)|^2 = \int_{-\infty}^{\infty} TF_x(t, f) df.$$

Substituting for the time-frequency spectral energy produces

$$|x(t)|^2 = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} A_x(t, \tau) e^{-i2\pi f\tau} d\tau \right\} df.$$

Rearranging the integrals yields

$$\begin{aligned} |x(t)|^2 &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} e^{-i2\pi f\tau} df \right\} A_x(t, \tau) d\tau \\ &= \int_{-\infty}^{\infty} \delta(\tau) A_x(t, \tau) d\tau \\ &= A_x(t, 0). \end{aligned}$$

Hence, the property is proved. †††

Property IV.11: A time-frequency spectral energy produces the power spectral energy as a marginal distribution with respect to time if the time-frequency autocorrelation function is defined with an ensemble average time-frequency expected value which uses a linear transformation depending on a time and a time difference.

Proof: The marginal distribution of the time-frequency spectral energy with respect to time is

$$|X(f)|^2 = \int_{-\infty}^{\infty} TF_x(t, f) dt.$$

Because the nonstationary ensemble average time-frequency expected value uses a linear transformation depending on a time and a time difference,

$$|X(f)|^2 = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} A_x(t, \tau) e^{-i2\pi f\tau} d\tau \right\} dt.$$

Since transformation (IV.47) implies $t = t_1$ and $\tau = t_1 - t_2$,

$$\begin{aligned}
|X(f)|^2 &= \int_{-\infty}^{\infty} \left\{ \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t_1) x_i^*(t_2) e^{-i2\pi f(t_1 - t_2)} dt_2 \right\} dt_1 \\
&= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left\{ \int_{-\infty}^{\infty} x_i(t_1) e^{-i2\pi f t_1} dt_1 \int_{-\infty}^{\infty} x_i^*(t_2) e^{-i2\pi f t_2} dt_2 \right\} \\
&= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \{X_i(f) X_i^*(f)\} = |X(f)|^2.
\end{aligned}$$

Hence, the property is proved. †††

Property IV.10 gives a simple method for testing whether a time-frequency spectral energy produces the instantaneous power. The test checks if $A_x(t, 0)$ equals the instantaneous power for all time. The following properties check whether the entries in Table II produce the instantaneous power.

Property IV.12: Using transformation (IV.47) or (IV.60) and a time average time-frequency expected value, the time-frequency spectral energy does not necessarily produce the instantaneous power.

Proof: First, transformation (IV.47) and time average time-frequency expected values imply

$$\begin{aligned}
A_x(t, \tau) &= E_A[x(t) x^*(t - \tau)] \\
&= \int_{-\infty}^{\infty} x(\xi) x^*(\xi - \tau) d\xi.
\end{aligned}$$

Evaluating $A_x(t, \tau)$ at $\tau = 0$ produces

$$\begin{aligned}
A_x(t, 0) &= \int_{-\infty}^{\infty} x(\xi) x^*(\xi) d\xi \\
&= \int_{-\infty}^{\infty} |x(\xi)|^2 d\xi \neq |x(t)|^2, \text{ for some } t.
\end{aligned}$$

The second transformation (IV.60) and time average time-frequency expected values imply

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t + \tau/2)x^*(t - \tau/2)] \\ &= \int_{-\infty}^{\infty} x(\xi + \tau/2)x^*(\xi - \tau/2)d\xi. \end{aligned}$$

Evaluating $A_x(t, \tau)$ at $\tau = 0$ produces

$$\begin{aligned} A_x(t, 0) &= \int_{-\infty}^{\infty} x(\xi)x^*(\xi)d\xi \\ &= \int_{-\infty}^{\infty} |x(\xi)|^2 d\xi \neq |x(t)|^2, \text{ for some } t. \end{aligned}$$

Because $A_x(t, 0) \neq |x(t)|^2$ for some time t , Property IV.10 states that the time-frequency spectral energy does not produce the instantaneous power. †††

Property IV.13: Using transformation (IV.47) or (IV.60) and either ensemble average time-frequency expected values or one sample function ensemble average time-frequency expected values, the time-frequency spectral energy produces the instantaneous power.

Proof: First, transformation (IV.47) and ensemble average time-frequency expected values imply

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t)x^*(t - \tau)] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i^*(t - \tau). \end{aligned}$$

Evaluating $A_x(t, \tau)$ at $\tau = 0$ produces

$$\begin{aligned} A_x(t, 0) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i^*(t) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N |x_i(t)|^2 = |x(t)|^2. \end{aligned}$$

The second transformation (IV.60) and ensemble average time-frequency expected values imply

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t + \tau/2)x^*(t - \tau/2)] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t + \tau/2)x_i^*(t - \tau/2). \end{aligned}$$

Evaluating $A_x(t, \tau)$ at $\tau = 0$ produces

$$\begin{aligned} A_x(t, 0) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i^*(t) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N |x_i(t)|^2 = |x(t)|^2. \end{aligned}$$

Because $A_x(t, 0) = |x(t)|^2$ for all time, Property IV.10 states that the time-frequency spectral energy produces the instantaneous power. †††

Property IV.12 indicates that time average time-frequency expected values do not allow the time-frequency spectral energy to produce the instantaneous power. This result implies that time average based time-frequency spectral energies do not provide useful time dependant features. However, since this time-frequency spectral energy corresponds to the spectrogram, the time varying spectral estimates are useful. Additionally, the time average time-frequency spectral energies' lack of producing the instantaneous power is expected because (II.103) and (II.104) indicate that the spectrogram produces the instantaneous power only for the impractical impulse window.

Property IV.13 implies that any time-frequency spectral energy based upon ensemble average time-frequency expected values produce the instantaneous power. Thus, the power spectral density, the Rihaczek distribution, and the Wigner distribution contain the instantaneous power as features. This result implies that ensemble average time-frequency expected value based time-frequency spectral energies are much more useful when using a time-frequency representations for extracting features. Property

IV.11 provides a method for checking whether a time-frequency spectral energy contains the spectral energy. However, Chapter III indicated that only features dependent on time are used regularly. Therefore, although Property IV.11 provides information on features that depend on frequency, determining which of the spectral energies in Table II contain the frequency dependent features is not provided.

The question that continues to linger is whether any technique exists which is capable of replacing the spectrogram for estimating time varying spectral energies. Because the periodogram is related to a spectrogram using a rectangular window, this question becomes if any time-frequency spectral energy exists that performs better than the periodogram. The periodogram's advantage is its real positive values, but the periodogram's lacking of the instantaneous power may hurt when analyzing highly nonstationary signals. Thus, does any real valued time-frequency spectral energy exist which allows computing the instantaneous power and has less anomalies than the Wigner distribution? Because no other reasonable time-frequency expected values seem to exist, defining different transformations is the only way to produce additional time-frequency spectral energies.

The search for a better time-frequency spectral energy begins by analyzing the ramifications of the following transformation:

$$\begin{array}{ccc} t = t_2 + k(t_2 - t_1) & \Leftrightarrow & t_1 = t + [k + 1]\tau \\ \tau = t_1 - t_2 & & t_2 = t + k\tau \end{array} \quad (\text{IV.74})$$

Transformation (IV.74) generalizes the two previous transformations since $k = -1$ produces transformation (IV.47) while $k = -1/2$ produces transformation (IV.60). Because transformation (IV.74) contains both previous transformations as special cases, properties based upon using (IV.74) adhere to the least restrictive versions of the previous properties. Thus, a time-frequency spectral energy based upon transformation (IV.74) and time average time-frequency expected values contain real values but do not allow computing

the instantaneous power. Conversely, in general, time-frequency spectral energies based upon transformation (IV.74) and either time average time-frequency expected values or one sample function time average time-frequency expected values contains complex values but allow computing the instantaneous power. The question is whether values of k exist which produce real time-frequency spectral energies and allow computing the instantaneous power. The following property answers this question.

Property IV.14: Using transformation (IV.74) and a one sample function ensemble average time-frequency expected value, $A_x(t, \tau) = A_x(t, -\tau)$ for all real $x(t)$ if and only if $k = -1/2$.

Proof: Since $A_x(t, \tau)$ uses transformation (IV.74) and a one sample function ensemble average time-frequency expected value,

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t + [k + 1]\tau)x^*(t + k\tau)] \\ &= E[x(t + [k + 1]\tau)x(t + k\tau)] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t + [k + 1]\tau)x_i(t + k\tau) \\ &= x(t + [k + 1]\tau)x(t + k\tau). \end{aligned}$$

Similarly,

$$\begin{aligned} A_x(t, -\tau) &= E[x(t - [k + 1]\tau)x(t - k\tau)] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t - [k + 1]\tau)x_i(t - k\tau) \\ &= x(t - [k + 1]\tau)x(t - k\tau). \end{aligned}$$

Suppose $A_x(t, \tau) = A_x(t, -\tau)$, then $x(t + [k + 1]\tau)x(t + k\tau) = x(t - [k + 1]\tau)x(t - k\tau)$.

If $k = 0$, then $x(t + \tau)x(t) = x(t - \tau)x(t)$ so that $x(t + \tau) = x(t - \tau)$. Thus, $k \neq 0$ since signals exists where $x(t + \tau) \neq x(t - \tau)$.

Now assume $k \neq 0$ so that either

$$x(t + [k + 1]\tau) = x(t - [k + 1]\tau) \text{ and } x(t + k\tau) = x(t - k\tau)$$

or

$$x(t + [k + 1]\tau) = x(t - k\tau) \text{ and } x(t - [k + 1]\tau) = x(t + k\tau).$$

For the first case, the only way $x(t + k\tau) = x(t - k\tau)$ is for $t + k\tau = t - k\tau$ which implies $k = -k$. Since $k = 0$ is the only solution of $k = -k$ and because $k \neq 0$, the first case is not possible.

Checking the second case begins by investigating when $x(t + [k + 1]\tau) = x(t - k\tau)$.

The only way for $x(t + [k + 1]\tau) = x(t - k\tau)$ is when $t + [k + 1]\tau = t - k\tau$ which implies $k = -1/2$. Continuing with the second case, the only way for $x(t - [k + 1]\tau) = x(t + k\tau)$ is when $t - [k + 1]\tau = t + k\tau$ which implies $k = -1/2$. Thus the second case is valid only when $k = -1/2$. Hence if $A_x(t, \tau) = A_x(t, -\tau)$ then $k = -1/2$.

Suppose $k = -1/2$. Then

$$A_x(t, \tau) = x(t + \tau/2)x(t - \tau/2) \text{ and}$$

$$A_x(t, -\tau) = x(t - \tau/2)x(t + \tau/2).$$

Thus, $A_x(t, \tau) = A_x(t, -\tau)$.

Hence, the property has been proved in both directions. †††

Property IV.14 indicates that, at least for real data, $k = -1/2$ produces the only linear transformation of this form which contains a real valued time-frequency spectral energy. Thus, transformation (IV.60) is the only linear transformation which allows both a real valued time-frequency spectral energy and instantaneous power calculation. This result reduces the possibilities for alternate time-frequency spectral energies because transformation (IV.74) is quite general. However, the possibility exists for defining nonlinear transformations.

The main problem with nonlinear transformations comes from their difficult mathematical structure. Nonetheless, the following nonlinear transformation is worth studying:

$$\begin{aligned} t = t_1 \\ |\tau| = t_1 - t_2 \end{aligned} \Leftrightarrow \begin{aligned} t_1 = t \\ t_2 = t - |\tau| \end{aligned} \quad (IV.75)$$

Although transformation (IV.75) appears relatively harmless, the absolute value function is not an analytic function, so the transformation is not differentiable. As such, transformation (IV.75) is difficult to analyze. However, transformation (IV.75), used in conjunction with the time-frequency expected value, produces time-frequency spectral energies which contain nice properties. The following properties indicate how the time-frequency spectral energies behaves when defined with transformation (IV.75).

Property IV.15: Using transformation (IV.75) and a time average time-frequency expected value, $A_x(t, \tau) = A_x(t, -\tau)$ for either complex or real $x(t)$.

Proof: Since $A_x(t, \tau)$ is defined by (IV.75) and a time average time-frequency expected value,

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t)x^*(t - |\tau|)] \\ &= \langle x(t)x^*(t - |\tau|) \rangle \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(\xi)x^*(\xi - |\tau|)d\xi. \end{aligned}$$

$$\begin{aligned} \text{Similarly, } A_x(t, -\tau) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(\xi)x^*(\xi - |-\tau|)d\xi \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(\xi)x^*(\xi - |\tau|)d\xi. \end{aligned}$$

Thus, $A_x(t, \tau) = A_x(t, -\tau)$ and the property is proved for both complex and real $x(t)$. †††

Property IV.16: Using transformation (IV.75) and an ensemble average time-frequency expected value, $A_x(t, \tau) = A_x(t, -\tau)$ for either complex or real $x(t)$.

Proof: Since $A_x(t, \tau)$ is defined by (IV.75) and an ensemble average time-frequency expected value,

$$\begin{aligned}
A_x(t, \tau) &= E_A[x(t)x^*(t - |\tau|)] \\
&= E[x(t)x^*(t - |\tau|)] \\
&= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i^*(t - |\tau|).
\end{aligned}$$

Similarly, $A_x(t, -\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i^*(t - |-\tau|)$

$$= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i^*(t - |\tau|).$$

Thus, $A_x(t, \tau) = A_x(t, -\tau)$ and the property is proved for both complex and real $x(t)$. †††

Property IV.17: The time-frequency spectral energy may not produce the instantaneous power when the time-frequency autocorrelation function is defined with transformation (IV.75) and a time average time-frequency expected value.

Proof: Using transformation (IV.75) and time average time-frequency expected values yields

$$\begin{aligned}
A_x(t, \tau) &= E_A[x(t)x^*(t - |\tau|)] \\
&= \int_{-\infty}^{\infty} x(\xi)x^*(\xi - |\tau|)d\xi.
\end{aligned}$$

Evaluating $A_x(t, \tau)$ at $\tau = 0$ produces

$$\begin{aligned}
A_x(t, 0) &= \int_{-\infty}^{\infty} x(\xi)x^*(\xi)d\xi \\
&= \int_{-\infty}^{\infty} |x(\xi)|^2 d\xi \neq |x(t)|^2, \text{ for some } t.
\end{aligned}$$

Because $A_x(t, \tau) \neq |x(t)|^2$ for some time, Property IV.10 states that the time-frequency spectral energy does not produce the instantaneous power. †††

Property IV.18: The time-frequency spectral energy produces the instantaneous power when the time-frequency autocorrelation function is defined with transformation (IV.75) and an ensemble average time-frequency expected value.

Proof: Using transformation (IV.75) and ensemble average time-frequency expected values yields

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t)x^*(t - |\tau|)] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i^*(t - |\tau|). \end{aligned}$$

Evaluating $A_x(t, \tau)$ at $\tau = 0$ produces

$$\begin{aligned} A_x(t, 0) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i^*(t) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N |x_i(t)|^2 = |x(t)|^2. \end{aligned}$$

Because $A_x(t, \tau) = |x(t)|^2$ for all time, Property IV.10 states that the time-frequency spectral energy produces the instantaneous power. †††

Properties IV.15 and IV.16 imply that regardless of whether a signal is real or complex valued, transformation (IV.75) and any of the defined time-frequency expected values produce real time-frequency spectral energies. Thus, results provided by Properties IV.15 and IV.16 actually are stronger than previously found with the other transformations. Notice that for complex $x(t)$, $A_x(t, \tau) = A_x(t, -\tau)$ replaces $A_x(t, \tau) = A_x^*(t, -\tau)$ as the main condition. The lacking of the complex conjugate function makes transformation (IV.75) slightly better when compared to transformations (IV.47) and (IV.48).

Property IV.17 indicates that the time average time-frequency expected value does not allow computing the instantaneous power. Alternately, Property IV.18 states that the ensemble average time-frequency expected value and the one sample function time-

frequency expected value based time-frequency spectral energy produces the instantaneous power. Thus, when considering the instantaneous power, transformation (IV.75) based time-frequency spectral energies behave identically with time-frequency spectral energies based upon transformations (IV.47) and (IV.60). Therefore, time-frequency spectral energies based upon transformation (IV.75) provide the same or better properties than those which use transformation (IV.47) and (IV.60). One remaining question is whether the three different time-frequency spectral energies produced by (IV.75) correspond to any currently known techniques? The immediately following text answers this question.

The first missing time-frequency spectral energy defines the time-frequency autocorrelation function with transformation (IV.75) and the time average time-frequency expected value. Thus,

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t)x^*(t - |\tau|)] \\ &= \int_{-\infty}^{\infty} x(\xi)x^*(\xi - |\tau|)d\xi. \end{aligned} \quad (IV.76)$$

Because $A_x(t, \tau)$ is similar to (IV.52) and is even with respect to τ , $A_x(t, \tau) = \langle R_x(\tau) \rangle$.

Therefore, by using (IV.51), the time-frequency spectral energy becomes

$$\begin{aligned} TF_x(t, f) &= \int_{-\infty}^{\infty} A_x(t, \tau)e^{-2\pi f\tau}d\tau \\ &= \frac{1}{T}|X(t)|^2, \end{aligned} \quad (IV.77)$$

where transformation (IV.75) and the time average time-frequency expected value define the time-frequency autocorrelation function. Thus, the time-frequency spectral energy corresponds to the periodogram.

The second missing time-frequency spectral energy defines the time-frequency autocorrelation function with transformation (IV.75) and the ensemble average time-frequency expected value. Thus,

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t)x^*(t - |\tau|)] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i(t)x_i^*(t - |\tau|). \end{aligned} \quad (\text{IV.78})$$

This time-frequency autocorrelation function is similar to the stationary autocorrelation function except for the absolute value. Because the absolute value's only purpose is to ensure that $A_x(t, \tau)$ is even with respect to τ , $A_x(t, \tau) = R_x(\tau)$ since $R_x(\tau)$ is an even function with respect to τ . Hence using (IV.55),

$$\begin{aligned} TF_x(t, f) &= \int_{-\infty}^{\infty} A_x(t, f) e^{-i2\pi f\tau} d\tau \\ &= S_x(\tau), \end{aligned} \quad (\text{IV.79})$$

where transformation (IV.75) and the ensemble average time-frequency expected value define the time-frequency autocorrelation function. Thus, the time-frequency spectral energy represents the power spectral density.

The third and last missing time-frequency spectral energy defines the time-frequency autocorrelation function with transformation (IV.75) and a one sample function ensemble average time-frequency expected value. Thus,

$$\begin{aligned} A_x(t, \tau) &= E_A[x(t)x^*(t - |\tau|)] \\ &= x(t)x^*(t - |\tau|). \end{aligned} \quad (\text{IV.80})$$

Substituting (IV.80) directly into the time-frequency spectral energy produces

$$\begin{aligned}
TF_x(t, f) &= \int_{-\infty}^{\infty} A_x(t, \tau) e^{-i2\pi f\tau} d\tau \\
&= \int_{-\infty}^{\infty} x(t)x^*(t - |\tau|) e^{-i2\pi f\tau} d\tau.
\end{aligned} \tag{IV.81}$$

Recall that Turner's instantaneous power spectrum, (II.58), is identical to this time-frequency spectral energy. Thus, using transformation (IV.75) and a one sample function ensemble average time-frequency expected value produces Turner's instantaneous power spectrum.

Since the three missing time-frequency spectral energies correspond to known quantities, these three time-frequency spectral energies can be compared to the previous time-frequency spectral energies. The comparison begins by augmenting Table II with the new results to produce Table IV. Table IV is interesting because it summarizes the entire chapter. Recall that transformation (IV.47) represents the most common method of converting the two correlation time variables into a time variable and a time difference variable. Transformation (IV.47) can be thought of as the natural first attempt in time-frequency spectral energies since the transformation defines the original periodogram and the original power spectral density. Transformation (IV.60) arose from the Wigner distribution although for ergodic or stationary random processes, transformation (IV.60) produces identical results as transformation (IV.47). Additionally, Property IV.14 showed that transformation (IV.60) is one of a few linear transformations which produce real valued time-frequency spectral energies. Because linear transformations are limited, the nonlinear transformation given in (IV.75) was investigated. Transformation (IV.75) can be thought of as a patch designed to force real valued time-frequency spectral energies using transformation (IV.47). Since transformation (IV.75) is nonlinear, proofs and analysis are difficult. In fact, although many other nonlinear transformations exist, they lack intuitive notions and become mathematically cumbersome. Thus, transformation

TABLE IV
A COMPARISON OF NINE TIME-FREQUENCY SPECTRAL ENERGIES

	Transformation IV.47	Transformation IV.60	Transformation IV.75
Time Average Time-Frequency Expected Value	Periodogram	Periodogram	Periodogram
Ensemble Average Time-Frequency Expected Value	Power Spectral Density	Power Spectral Density	Power Spectral Density
One Sample Function Ensemble Average Time-Frequency Expected Value	Rihaczek Distribution	Wigner Distribution	Turner's Instantaneous Power Spectrum

(IV.75) is the only nonlinear transformation considered.

When investigating Table IV, the fact that all three transformations produce both the periodogram and the power spectral density (for ergodic and stationary random processes) is amazing. Even for nonstationary random processes, the three transformations produce interesting time-frequency spectral energies for time average and ensemble average time-frequency expected values. However, since transformations (IV.60) and (IV.75) produce real time-frequency spectral energies, they are preferable. Additionally, transformation (IV.75) is preferred over (IV.60) because the time-frequency autocorrelation function for transformation (IV.75) is an even function of τ for both real and complex signals. Although using transformation (IV.60) in signal processing procedures is not performed, computing time-frequency autocorrelation functions with transformation (IV.60) is actually the most computationally efficient of the three transformations. The reader should note that no matter what transformation is used, both the time average and the ensemble average time-frequency expected value corresponds to widely implemented techniques. Thus, using any of the six methods in the top two rows in Table IV produces reasonable results.

The three techniques on the bottom row of Table IV are probably the most interesting of the entire group. Instead of producing similar time-frequency spectral energies, one sample function time-frequency expected value based time-frequency autocorrelation functions produce widely different time-frequency spectral energies for each transformation. The Rihaczek distribution, generated using transformation (IV.47), is not widely used in practice because it contains complex values. Complex valued time-frequency spectral energies definitely cannot be used as spectral energy estimates for a particular time. However, the Rihaczek distribution contains the instantaneous power that indicates some nonstationary spectral energy information may exist. Because one goal is to replace the spectrogram, the Rihaczek distribution is no longer considered.

The Wigner distribution was rigorously discussed in Chapter III. Chapter III indicated that the Wigner distribution may contain nonstationary spectral information but in the form of features. Crossterm components and non-negative values prohibit replacing the spectrogram with the Wigner distribution. The question is whether Turner's instantaneous power spectrum could replace the spectrogram. The following property helps make the decision.

Property IV.19: Using transformation (IV.75) and a one sample function ensemble average time-frequency expected value, the time-frequency spectral energy produces crossterm components.

Proof: Let $x(t) = x_1(t) + x_2(t)$. Then

$$\begin{aligned} TF_x(t, f) &= \int_{-\infty}^{\infty} x(t)x^*(t - |\tau|)e^{-i2\pi f\tau} d\tau \\ &= \int_{-\infty}^{\infty} (x_1(t) + x_2(t))(x_1(t - |\tau|) + x_2(t - |\tau|))^* e^{-i2\pi f\tau} d\tau. \end{aligned}$$

Expanding the time-frequency spectral energy produces

$$\begin{aligned} TF_x(t, f) &= \int_{-\infty}^{\infty} x_1(t)x_1^*(t - |\tau|)e^{-i2\pi f\tau} d\tau + \int_{-\infty}^{\infty} x_2(t)x_2^*(t - |\tau|)e^{-i2\pi f\tau} d\tau \\ &\quad + \int_{-\infty}^{\infty} x_1(t)x_2^*(t - |\tau|)e^{-i2\pi f\tau} d\tau + \int_{-\infty}^{\infty} x_2(t)x_1^*(t - |\tau|)e^{-i2\pi f\tau} d\tau. \end{aligned}$$

Using the notation where the time-frequency spectral energy using $E_A[\xi, \alpha]$ is $TF_{\xi, \alpha}(t, f)$

implies

$$TF_x(t, f) = TF_{x_1}(t, f) + TF_{x_2}(t, f) + TF_{x_1, x_2}(t, f) + TF_{x_2, x_1}(t, f).$$

Because extra terms $TF_{x_1, x_2}(t, f)$ and $TF_{x_2, x_1}(t, f)$ are, in general, non-zero, crossterm com-

ponents exist. †††

Property IV.19 implies that the time-frequency spectral energy generated by transformation (IV.75) contains crossterm components. Thus, Turner's instantaneous power spectrum is no better at replacing the spectrogram than the Wigner distribution. Hence, one thesis conclusion is that no obvious method exists to date that allows estimating spectral energy in a reasonable fashion. Since Turner's instantaneous power spectrum may contain useful features, the Wigner distribution and Turner's instantaneous power spectrum are similar in their potential usefulness.

Two final thoughts concerning Table IV require a brief discussion before presenting the final conclusion. First, the bottom row time-frequency spectral energies all correspond to instantaneous power spectra. Instantaneous power spectra correspond to time-frequency representations which indicate how spectral energy changes over time. Thus, plotting the Rihaczek distribution, the Wigner distribution, or Turner's instantaneous power spectrum with respect to frequency does not appear like the spectrogram but like the spectrogram's derivative. The fact that the bottom three time-frequency spectral energies are instantaneous power spectra comes from the fact that all three techniques can produce the instantaneous power. Second, recall that the time-frequency spectral energies defined on the bottom row all use one sample function ensemble average time-frequency expected values. Because the one sample function ensemble average time-frequency expected value is a very poor ensemble average estimator (because the average contains only one term), the noisy quality of these time-frequency spectral energies may be directly attributed directly to the poor ensemble average estimates. Thus, no matter what transformation is defined, the one sample function time average time-frequency expected value may always produce time-frequency spectral energies that cannot replace the spectrogram for estimating time varying spectral energies.

Hence, it seems that producing the instantaneous power and estimating reasonable spectral energies are mutually exclusive operations. Thus, estimating spectral energy is left to the spectrogram. However, because the Rihaczek distribution, the Wigner distri-

bution, and Turner's instantaneous power spectrum may contain nonstationary spectral energy information in the form of the instantaneous power, any of the three time-frequency spectral energies may be useful for generating signal processing features. For comparative purposes, the Wigner distribution will be used to generate all future nonstationary spectral energy features.

Chapter Summary

This chapter has introduced the correlation approach to time-frequency representations. In defining the correlation approach, three new quantities, the time-frequency autocorrelation function, the time-frequency expected value, and the time-frequency spectral energy, were introduced. The time-frequency autocorrelation function is similar to an ordinary nonstationary autocorrelation function except that the time-frequency expected value replaces the normal expected value. The time-frequency expected value is closely related to the expected value except no implicit assumption is made concerning the underlying random process. As such, the time-frequency expected value can be defined in many different ways.

In order to place the most common nonstationary spectral energy techniques into the correlation approach framework, the time-frequency autocorrelation function, a function of two time variables, is written in terms of a time variable and a time difference variable. A linear or nonlinear transformation describes the change from the two time variables to a time variable and a time difference variable. Using the time-frequency autocorrelation function as a base, the time-frequency spectral energy is defined as the Fourier transformation of the time-frequency autocorrelation function with respect to the time difference. The time-frequency spectral energy is just a time-frequency representation based upon a pre-defined time difference dependent autocorrelation function.

Through an extensive number of properties, the time-frequency spectral energy was shown quite capable of producing the main time-frequency representations discussed in Chapter II along with techniques commonly used in signal processing. The correlation approach thus provides a general framework for studying different time-frequency representations. This general framework provides a similar forum that Cohen's generalized time-frequency representation gives except that the correlation approach is easier to understand and simpler to analyze. Engineers should understand time-frequency representations better when using the correlation approach because engineers are quite comfortable with autocorrelation functions. The properties proven in this chapter show how easily analyzing different time-frequency representations is when using the correlation approach. One last advantage of the correlation approach is that new time-frequency representations can easily be created by simply defining new nonlinear transformations or by defining new time-frequency expected values.

Near the end of this chapter, the correlation approach showed that requiring a time-frequency spectral energy to estimate spectral energies and produce spectral energy features is mutually exclusive. Therefore, as of now, it does not seem likely that estimating time varying spectral energies will be performed using alternate time-frequency representations. However, this chapter and Chapter III indicate that nonstationary spectral energy information may exist in some time-frequency spectral energies. The nonstationary spectral energy information is used in the next chapter to formulate a technique which attempts to indicate stationary and nonstationary portions of signals.

CHAPTER V

THE NONSTATIONARY INDICATOR

Introduction

Chapter IV indicated that time-frequency representations probably will not replace the spectrogram for estimating time varying spectral energies. However, both Chapters III and IV indicated that time-frequency distribution time-frequency spectral energies such as the Wigner distribution for example, may produce features or signals that contain some time varying spectral energy information. This chapter begins by reviewing the time varying spectral energy features already discussed in previous chapters. Concurrently, some new features, which many contain additional spectral energy information, are introduced. The proposed time-frequency spectral energy based features are tested on a new signal processing application. This signal processing application, called the nonstationary indicator, determines whether a signal is stationary or nonstationary.

The nonstationary indicator is a function whose input is any complex or real signal and whose output indicates if a signal is stationary or nonstationary. Because the time-frequency spectral energy features may provide time varying spectral energy information, these features form the basis of the nonstationary indicator. However, some mechanism for mapping the time-frequency spectral energy features to the nonstationary output is needed. Because the multilayer perceptron performs arbitrary mappings, the multilayer perceptron implements the nonstationary indicator's mapping.

Because the time-frequency spectral energy features base the nonstationary indicator, the feature overview is given before the nonstationary indicator description. The feature overview is based upon computing features on the Wigner distribution, but any of the one sample function time-frequency expected value based time-frequency spectral energies could have been used. Next, the nonstationary indicator and then the multilayer perceptron (which implements the nonstationary indicator) is covered in depth. Additionally, the Volterra expansion multilayer perceptron, which extends the basic multilayer perceptron, is discussed. Finally, the time-frequency spectral energy features and the Volterra expansion multilayer perceptron are tested.

Wigner Distribution Based Features

Recall again that the Wigner distribution is defined by

$$W_x(t, f) = \int_{-\infty}^{\infty} x(t + \tau/2) x^*(t - \tau/2) e^{-i2\pi f\tau} d\tau. \quad (\text{V.1})$$

The Wigner distribution was designed specifically to produce valid marginal densities. As such, the Wigner distribution is thought of as a two-dimensional probability density function (pdf) which indicates how spectral energy distributes as a function of time and frequency. Recall from Chapter II, at the beginning of the Wigner distribution discussion, that the Wigner distribution is actually within a multiplication factor from a pdf since the total energy may not equal 1. Considering the Wigner distribution as a two-dimensional pdf allows computing statistical quantities on the Wigner distribution. Although many statistical quantities could be computed, two-dimensional pdf moments seem capable of providing the most useful time varying spectral energy features. However, because the Wigner distribution is not a true pdf (recall from Chapter III that the Wigner distribution can be negative), that moments must be carefully defined. The following considers only the time dependent moments since the nonstationary indicator

requires features dependent on time.

Moment Based Features

In order to fully describe Wigner distribution based moments, the general theory of pdf moments must be explained. For any two-dimensional pdf, $f(x, y)$, one marginal density is defined by

$$f(y) = \int_{-\infty}^{\infty} f(x, y) dx. \quad (V.2)$$

Naturally the other marginal density, $f(x)$, is identical to (V.2) except that the integration is with respect to y . Note that integrating the marginal density $f(y)$ with respect to y will always equal 1 since $f(y)$ is a valid one-dimensional pdf. Because $f(y)$ is a pdf, expected values or moments can be computed. The generalized k^{th} -order moment of $f(y)$ is defined by

$$m_y^k = \int_{-\infty}^{\infty} y^k f(y) dy, \quad (V.3)$$

where k represents the moment order. In order to get this general theory in terms of the Wigner distribution, (V.3) is rewritten by substituting (V.2) in for $f(y)$ to produce

$$\begin{aligned} m_y^k &= \int_{-\infty}^{\infty} y^k \left\{ \int_{-\infty}^{\infty} f(x, y) dx \right\} dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^k f(x, y) dx dy. \end{aligned} \quad (V.4)$$

Notice that the marginal density, $f(y)$, directly depends on y , but the generalized k^{th} -order moment, m_y^k , has had y integrated out. Because moments that depend on time are desired

for the Wigner distribution, moments which depend on y are desired from $f(x, y)$.

Obtaining time dependent moments requires defining conditional moments. The generalized k^{th} -order conditional moment of $f(y)$ is

$$m_y^k(x) = \int_{-\infty}^{\infty} y^k f(y | x) dy, \quad (\text{V.5})$$

where $f(y | x)$ represents the conditional marginal density of y given a value of x . Using Bayes' rule [Pap84], (V.5) simplifies to

$$\begin{aligned} m_y^k(x) &= \int_{-\infty}^{\infty} y^k \frac{f(x, y)}{f(x)} dy \\ &= \frac{1}{f(x)} \int_{-\infty}^{\infty} y^k f(x, y) dy, \end{aligned} \quad (\text{V.6})$$

where $f(x, y)$ is just a two-dimensional pdf and $f(x)$ is the marginal density with respect to x .

The generalized k^{th} -order moment, (V.4), is different from the generalized conditional k^{th} -order moment, (V.6). To see a difference between the two moments, let $k = 1$. Substituting $k = 1$ into (V.4) and (V.6) produces

$$\begin{aligned} m_y^1 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy = \mu_y \\ m_y^1(x) &= \frac{1}{f(x)} \int_{-\infty}^{\infty} y f(x, y) dy = \mu_y(x). \end{aligned} \quad (\text{V.7})$$

Now μ_y is the mean y value over the entire two-dimensional pdf while $\mu_y(x)$ is the mean y value when x is set. These two means are different and require different interpretations. However, both means are widely used.

The generalized moments given above allow computing the average values of many order moments by changing the values of k . However, studying how values vary about the mean value is sometimes more interesting. Thus, the generalized moments are modified to produce central (tendency) moments. For $k > 1$, the generalized central k^{th} -order moment is

$$mc_y^k = \int_{-\infty}^{\infty} (y - m_y^1)^k f(y) dy, \quad (\text{V.8})$$

where m_y^1 is the mean value described in (V.7). Because mc_y^k represents how the variable y varies about x , there is no sense in studying the moments for $k < 2$. The generalized conditional central k^{th} -order moment is similarly defined by

$$mc_y^k(x) = \frac{1}{f(x)} \int_{-\infty}^{\infty} (y - m_y^1(x))^k f(x, y) dy, \quad (\text{V.9})$$

where $m_y^1(x)$ is the conditional mean described in (V.7) and from here on $k > 1$ in this context.

Although (V.8) and (V.9) define an infinite number of central moments, $k = 2$ produces the most commonly used central moment. When $k = 2$, (V.8) is called the variance, and (V.9) is called the conditional variance. Thus, the marginal density, the mean value, and the variance comprise the main statistical quantities. Why are higher order moments not often used? The reason stems from the fact that the mean and the variance completely describe the Gaussian pdf. Thus, higher order moments provide no additional information for Gaussian distributed random variables. Also, since Gaussian pdfs widely occur in theory and practice, many researchers ignore higher order moments. However, for non-Gaussian random variables, higher order moments provide additional statistical information than is contained in the mean and variance.

The first high order moment considered is obtained from the generalized central moment with $k = 3$. Substituting $k = 3$ into (V.8) and (V.9) yields

$$mc_y^3 = \int_{-\infty}^{\infty} (y - m_y^1)^3 f(y) dy$$

and

$$mc_y^3(x) = \frac{1}{f(x)} \int_{-\infty}^{\infty} (y - m_y^1(x))^3 f(x, y) dy. \quad (V.10)$$

Both mc_y^3 and $mc_y^3(x)$ can contain very large magnitudes, and thus are typically normalized. This normalization, called the skewness, is defined as

$$SK_y = \int_{-\infty}^{\infty} \left(\frac{y - m_y^1}{\sqrt{m_y^2}} \right)^3 f(y) dy, \quad (V.11)$$

where m_y^2 is the variance. Similarly, the conditional skewness normalizes to

$$SK_y(x) = \frac{1}{f(x)} \int_{-\infty}^{\infty} \left(\frac{y - m_y^1(x)}{\sqrt{m_y^2(x)}} \right)^3 f(x, y) dy, \quad (V.12)$$

where $m_y^2(x)$ is the conditional variance. Both skewness equations are dimensionless quantities that correspond to a pdf's asymmetry about the mean. Thus, a skewness value of 0 indicates a pdf contains perfectly symmetry about its mean. As an example, the Gaussian distribution's skewness value equals 0.

Another high order moment is obtained from the generalized central moment by setting $k = 4$. Substituting $k = 4$ into (V.8) and (V.9) produces

$$mc_y^4 = \int_{-\infty}^{\infty} (y - m_y^1)^4 f(y) dy$$

and

$$mc_y^4(x) = \frac{1}{f(x)} \int_{-\infty}^{\infty} (y - m_y^1(x))^4 f(x, y) dy. \quad (V.13)$$

As with the third order moments, mc_y^4 and $mc_y^4(x)$ both can be large and therefore, the fourth order moments are normalized. This normalization, called the kurtosis, is defined as

$$KU_y = \int_{-\infty}^{\infty} \left(\frac{y - m_y^1}{\sqrt{m_y^2}} \right)^4 f(y) dy - 3 \quad (\text{V.14})$$

while the conditional kurtosis is defined

$$KU_y(x) = \frac{1}{f(x)} \int_{-\infty}^{\infty} \left(\frac{y - m_y^1(x)}{\sqrt{m_y^2(x)}} \right)^4 f(x, y) dy - 3 \quad (\text{V.15})$$

The kurtosis definitions are dimensionless quantities that correspond to the relative peakedness or flatness of the underlying pdf compared to the Gaussian distribution's peakedness. Positive kurtosis values imply a flat distribution while negative kurtosis values imply a peak like distribution. As an example, the Gaussian distribution's kurtosis equals 0.0. Thus, the factor of 3 in (V.14) and (V.15) shifts the kurtosis so that the peakedness and flatness corresponds directly to the Gaussian distribution.

Even larger higher order moments could be defined, but higher order moments typically stop with $k = 4$. Thus, the general moment overview has described the marginal density, the mean value, the variance, the skewness, and the kurtosis in terms of true pdfs. The next step adapts these five statistical quantities to the Wigner distribution. Additionally, the necessity of including the skewness and kurtosis in the Wigner distribution requires investigation.

Converting the pdf based statistical quantities to the Wigner distribution requires care since the Wigner distribution really is not a pdf. The Wigner distribution is not a true pdf because the Wigner distribution can contain negative values (recall from Chapter III). In addition, the double integral of the Wigner distribution produces the total signal energy instead of 1 which a true pdf would produce. However, Chapter III already stated that the Wigner distribution marginal density and mean value correspond to the instant-

neous power and the instantaneous frequency, respectively. Thus, converting the other statistical quantities to the Wigner distribution may also provide additional time varying spectral energy information.

Converting statistical quantities to the Wigner distribution begins by comparing notation. First, the pdf $f(x, y)$ used in the general moment discussion is substituted with the Wigner distribution, $W_x(t, f)$. Because only features dependent on time are desired, variable x in the general moment discussion now represents time t . Thus, the marginal density (V.2) becomes

$$IP_w(t) = |x(t)|^2 = \int_{-\infty}^{\infty} W_x(t, f) df, \quad (V.16)$$

where $|x(t)|^2$ is the instantaneous power and $IP_w(t)$ signifies that the instantaneous power was computed using Wigner distribution. Since $W_x(t, f)$ is not a true two-dimensional pdf, $IP_w(t)$ is not a true one-dimensional pdf.

The next converted statistical value is the conditional mean. Substituting the appropriate Wigner distribution quantities, the conditional mean defined by (V.6) becomes

$$IF_w(t) = m_f^1(t) = \frac{1}{IP_w(t)} \int_{-\infty}^{\infty} f W_x(t, f) df, \quad (V.17)$$

where $IF_w(t)$ represents the instantaneous frequency. Note that (V.17) is defined only when $IP_w(t) \neq 0$. The Wigner distribution conditional mean (the instantaneous frequency) produces the centroid frequency at a given time. Chapter III shows an example that indicates the instantaneous frequency may provide useful time varying spectral energy information.

The conditional variance is the next statistical quantity converted to the Wigner distribution. Substituting the proper variables, the conditional variance (V.9) with $k = 2$ becomes

$$VA_w(t) = mc_f^2(t) = \frac{1}{IP_w(t)} \int_{-\infty}^{\infty} (f - IF_w(t))^2 W_x(t, f) df, \quad (V.18)$$

where $VA_w(t)$ corresponds to the Wigner distribution variance. As with the Wigner distribution instantaneous frequency, the Wigner distribution variance is valid only when $IP_w(t) \neq 0$. Additionally, even though the term variance is used, variance properties such as non-negativeness cannot be applied to the Wigner distribution variance. The reason that the Wigner distribution variance does not act like a normal variance is that the Wigner distribution contains negative values, whereas pdf's do not. Thus, interpreting the Wigner distribution variance must be carefully done.

Because the mean and variance completely describe a Gaussian distribution, is it necessary to define any more Wigner distribution moments? The only way to answer this question is by estimating how Wigner distribution values distribute. Current and past research on the Wigner distribution does not cover any theoretical or any experimental aspects of how Wigner distribution coefficients distribute. Thus, a technique for estimating the Wigner distribution coefficient distribution is required.

First, a sequence of zero mean and unit variance Gaussian random numbers make up a signal. Random numbers are used to prevent bias in the distribution estimates. The Wigner distribution of the random signal is computed using a Gaussian window. The histogram of the Wigner distribution coefficients for all times and all frequencies is displayed in Figure V.1. In order to compare the Wigner distribution coefficient histogram to existing histograms, the mean and the variance of the Wigner distribution coefficients was computed. Next, using the Wigner distribution coefficient mean and variance, a Gaussian histogram and a Laplacian histogram were computed and displayed in Figure V.1. Notice that the Wigner distribution coefficient histogram appears much more like

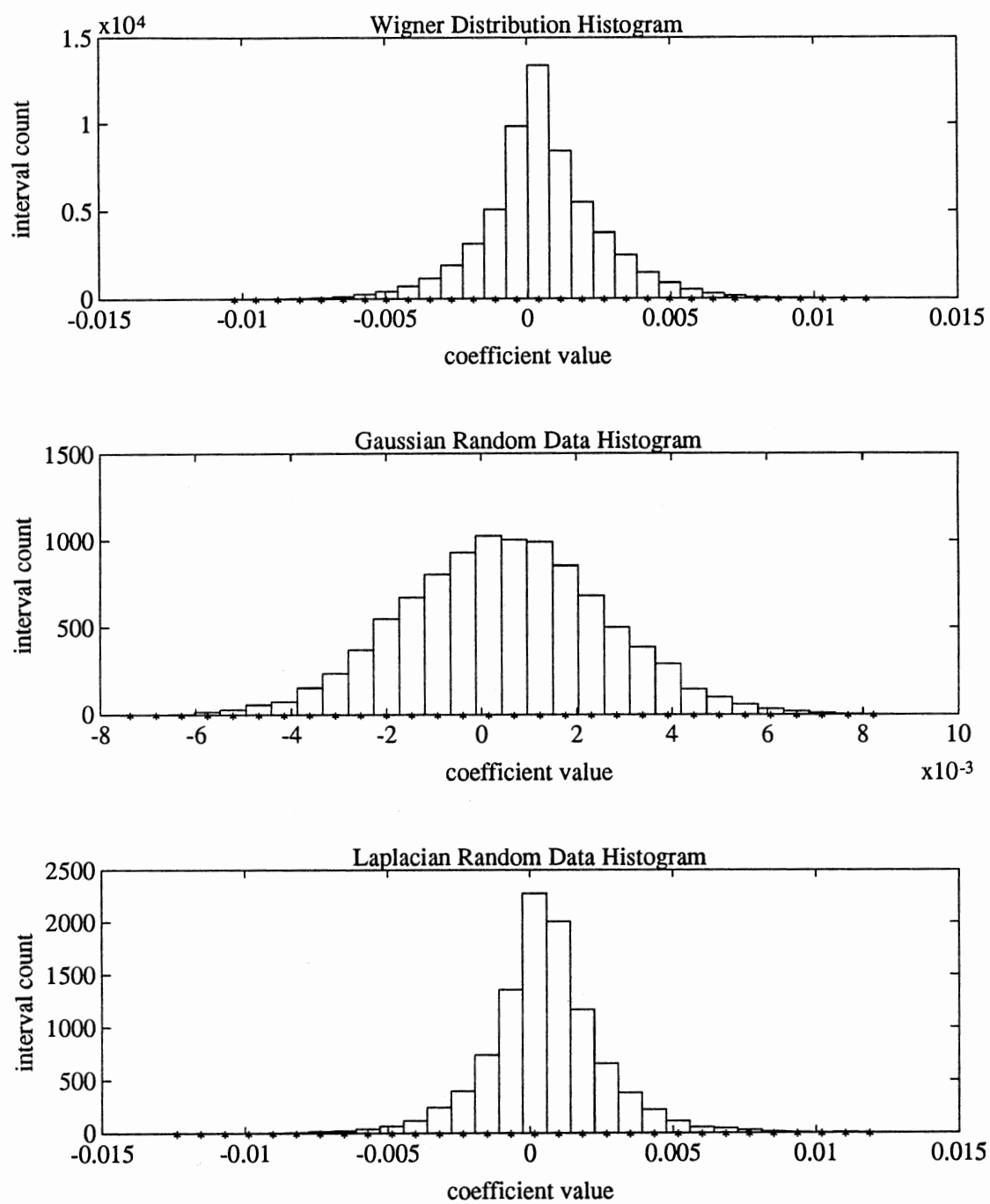


Figure V.1 Histograms of the Wigner, Gaussian, and Laplacian Distributions

the Laplacian histogram as opposed to the Gaussian histogram. Thus, Wigner distribution coefficients are not Gaussian distributed. Hence, higher order moments provide additional information not available in the mean and variance.

The first higher order moment placed into the context of the Wigner distribution is the conditional skewness. The conditional skewness converts to the Wigner distribution skewness by appropriately replacing variables in (V.12) to produce

$$SK_w(t) = SK_y(t) = \frac{1}{IP_w(t)} \int_{-\infty}^{\infty} \left(\frac{f - IF_w(t)}{\sqrt{|VA_w(t)|}} \right)^3 W_x(t, f) df, \quad (V.19)$$

where $SK_w(t)$ represents the Wigner distribution skewness. Notice that an absolute value was included in (V.12). The absolute value prevents having to take the square root of negative Wigner distribution variances. Also, (V.19) is valid only when both the Wigner distribution instantaneous power and variance are nonzero.

The last converted statistical quantity is the conditional kurtosis. Substituting appropriately, the conditional kurtosis, (V.15), becomes

$$KU_w(t) = KU_y(t) = \frac{1}{IP_w(t)} \int_{-\infty}^{\infty} \left(\frac{f - IF_w(t)}{\sqrt{|VA_w(t)|}} \right)^4 W_x(t, f) df, \quad (V.20)$$

where $KU_w(t)$ corresponds to the Wigner distribution kurtosis. Once again, (V.20) is valid only for times t when the Wigner distribution instantaneous power and variance are nonzero. Because the kurtosis is difficult to interpret, the Wigner distribution kurtosis is even more difficult to interpret. However, because the Wigner distribution coefficients are non-Gaussian, even the Wigner distribution kurtosis may provide some additional time varying spectral information.

As with the generalized moments and central moments, additional Wigner distribution moments could be defined. However, because higher order moments are difficult to interpret, the number of Wigner distribution moments was kept at the same number.

Thus, the five Wigner distribution statistical quantities represent the Wigner distribution features which implement the nonstationary indicator. Thus, the Wigner distribution has basically been reduced to five values for each time. The nonstationary indicator is described next.

The Nonstationary Indicator

Because Chapter IV answered the question of whether time-frequency spectral energies can replace the spectrogram for estimating time varying spectral energies, only one of the two main thesis goals is left. The unanswered thesis goal determines whether time-frequency spectral energies produce information that allow implementing signal processing algorithms which have never been possible or which are difficult using current techniques. The next paragraph introduces a signal processing application which has not been successfully implemented. This new signal processing algorithm's success depends on the fact that time-frequency spectral energies contain time varying signal information that does not currently exist.

The new time-frequency spectral energy based signal processing application is called the nonstationary indicator. The nonstationary indicator is a function which depends on a time-frequency spectral energy and whose single output indicates whether the signal at time t classifies as a stationary or nonstationary signal. Mathematically, the nonstationary indicator is defined as

$$NI(TF_x(t, f), t) \in [-1, 1] \in \mathfrak{R}$$

and

$$NI(TF_x(t, f), t) \stackrel{\Delta}{=} \begin{cases} 1, & \text{if } x(t) \text{ is stationary} \\ -1, & \text{if } x(t) \text{ is nonstationary} \end{cases}, \quad (\text{V.21})$$

where $NI(TF_x(t, f), t)$ represents the nonstationary indicator and explicitly shows the dependence on a time-frequency spectral energy and time. Because this chapter is only

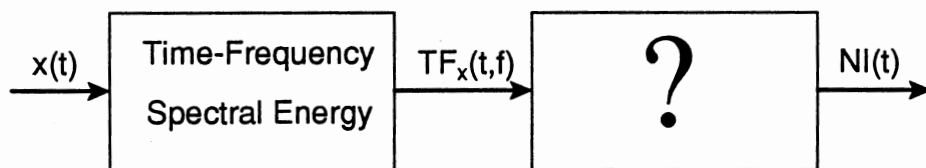


Figure V.2 The Basic Nonstationary Indicator System

concerned with applying time-frequency spectral energies to signal processing, the explicit nonstationary indicator dependence on a time-frequency spectral energy is assumed so that

$$NI(t) = NI(TF_x(t, f), t). \quad (V.22)$$

Thus, the nonstationary indicator is a time dependent real valued function whose values are bounded between -1 and 1 , inclusive. Notice that the nonstationary indicator definition does not state what values should occur for signals that are neither totally stationary nor totally nonstationary. Additionally, the nonstationary indicator definition says nothing about how the function should be implemented except that a time-frequency spectral energy should form the basis of the technique. Figure V.2 illustrates the basic nonstationary indicator definition. The question mark in the nonstationary indicator function block signifies the fact that no implementation method has been designated yet.

The nonstationary indicator's dependence on a time-frequency spectral energy allows many of possible implementation schemes. However, because the goal of any signal processing technique is for real time applications, deciding how to implement the

nonstationary indicator should consider the amount of processing time. As such, computing the nonstationary indicator directly on the two-dimensional time-frequency spectral energy may not be practical since at any time, $TF_x(t, f)$, could contain as many as 512 data points. Similar problems occur quite often in image processing when algorithms designed to compute directly on an image require too much computation time when implementing an image processing algorithm in a real world system. To circumvent the executing time problem, many signal processing algorithms are based on features instead of on the actual data [Tau74]. Because features contain significantly less data than the original representation, using features in lieu of the original representation reduces the computation time. Additionally, if the selected features contain good signal information, then the overall signal processing system can perform quite robustly. Therefore, features based upon time-frequency spectral energies implement the nonstationary indicator.

Because the nonstationary indicator requires information concerning whether a signal is stationary or nonstationary, time-frequency spectral energy features should also contain this information. Recall that a nonstationary system can be described as a system whose spectral energy changes with time. Thus, features which provide information concerning changing spectral energy could possibly identify stationary and nonstationary signals. Since the nonstationary indicator depends on time, the five time dependent statistical based features, described in the previous section, might provide the proper information. Therefore, the instantaneous power ($IP_w(t)$), the instantaneous frequency ($IF_w(t)$), the variance ($VA_w(t)$), the skewness ($SK_w(t)$), and the kurtosis ($KU_w(t)$) will be used as the time-frequency spectral energy features that form the basis of the nonstationary indicator. Because these five moment based features were defined by the Wigner distribution, the Wigner distribution is assumed to implement the general time-frequency spectral energy. Figure V.3 illustrates how the basic nonstationary indicator system,

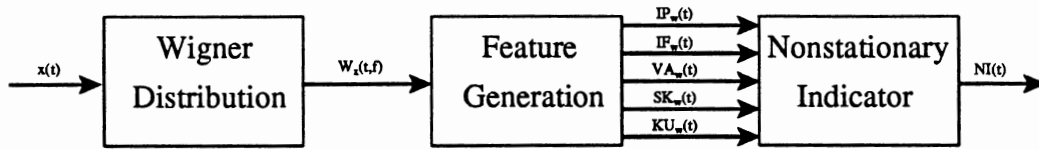


Figure V.3 The Signal Processing System Which Generates the Nonstationary Indicator

described in Figure V.2, is modified to include the use of Wigner distribution features. Note that nothing has yet been written as to how the features produce the nonstationary indicator.

Before deciding how Wigner distribution features transform into the nonstationary indicator, a few examples are presented to show the nonstationary indicator's ideal performance. Figure V.4 presents the single sinusoid (I.8) and the corresponding ideal nonstationary indicator. Because the sinusoid does not change frequency over time, the signal is stationary, and the nonstationary indicator equals 1. Figure V.5 displays a chirp signal (I.9) and its ideal nonstationary indicator. Since a chirp signal changes frequency over time, the signal is nonstationary and the nonstationary indicator equals -1. Finally, Figure V.6 presents a signal containing three concatenated sinusoids (I.11) and the signal's ideal nonstationary indicator. Because each of the three sinusoids does not change over time, the ideal nonstationary indicator equals 1 during each sinusoid. During transition regions, where the frequency of the sinusoid changes, the ideal nonstationary

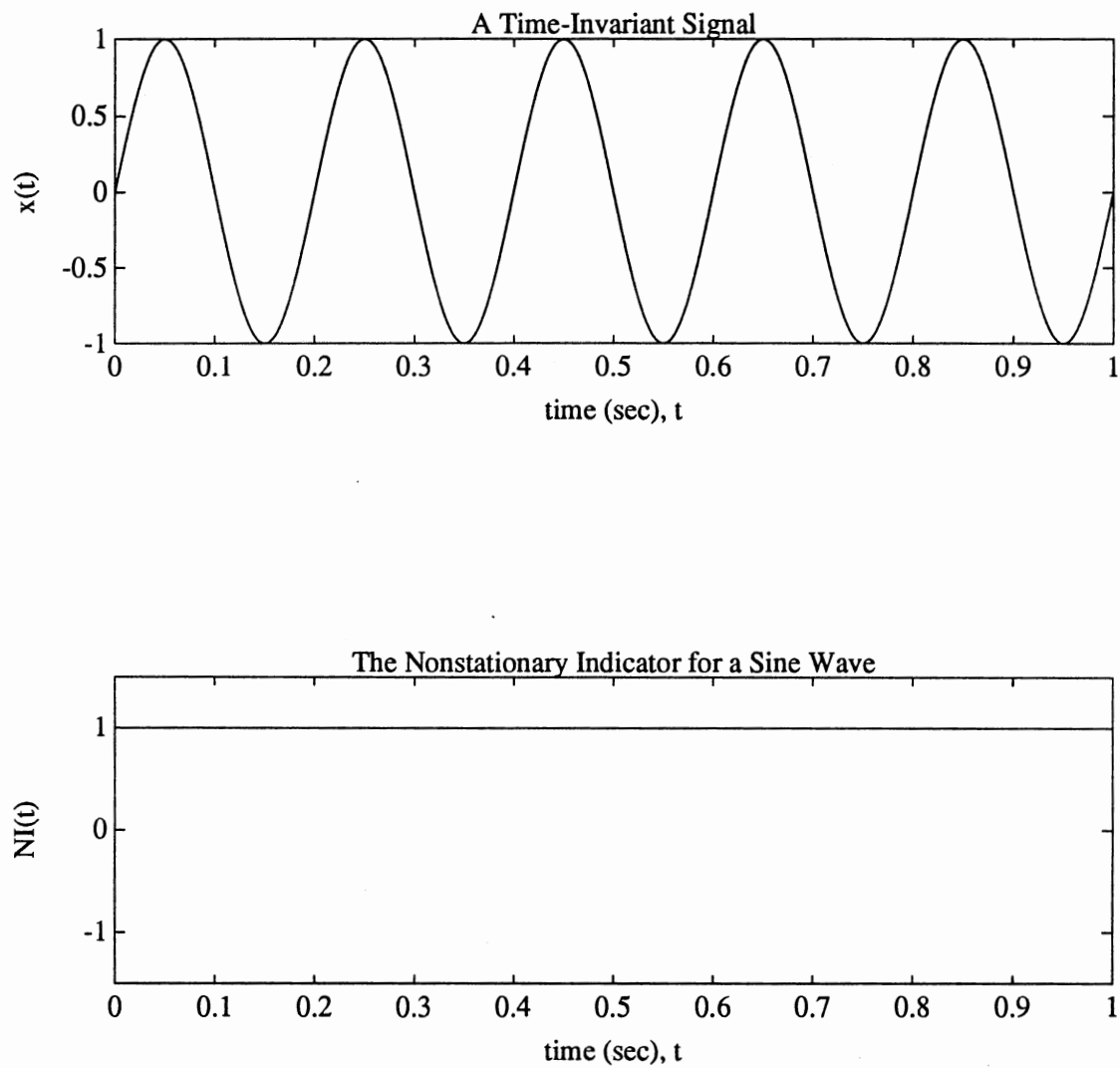


Figure V.4 The Ideal Nonstationary Indicator Performance for a Sinusoid

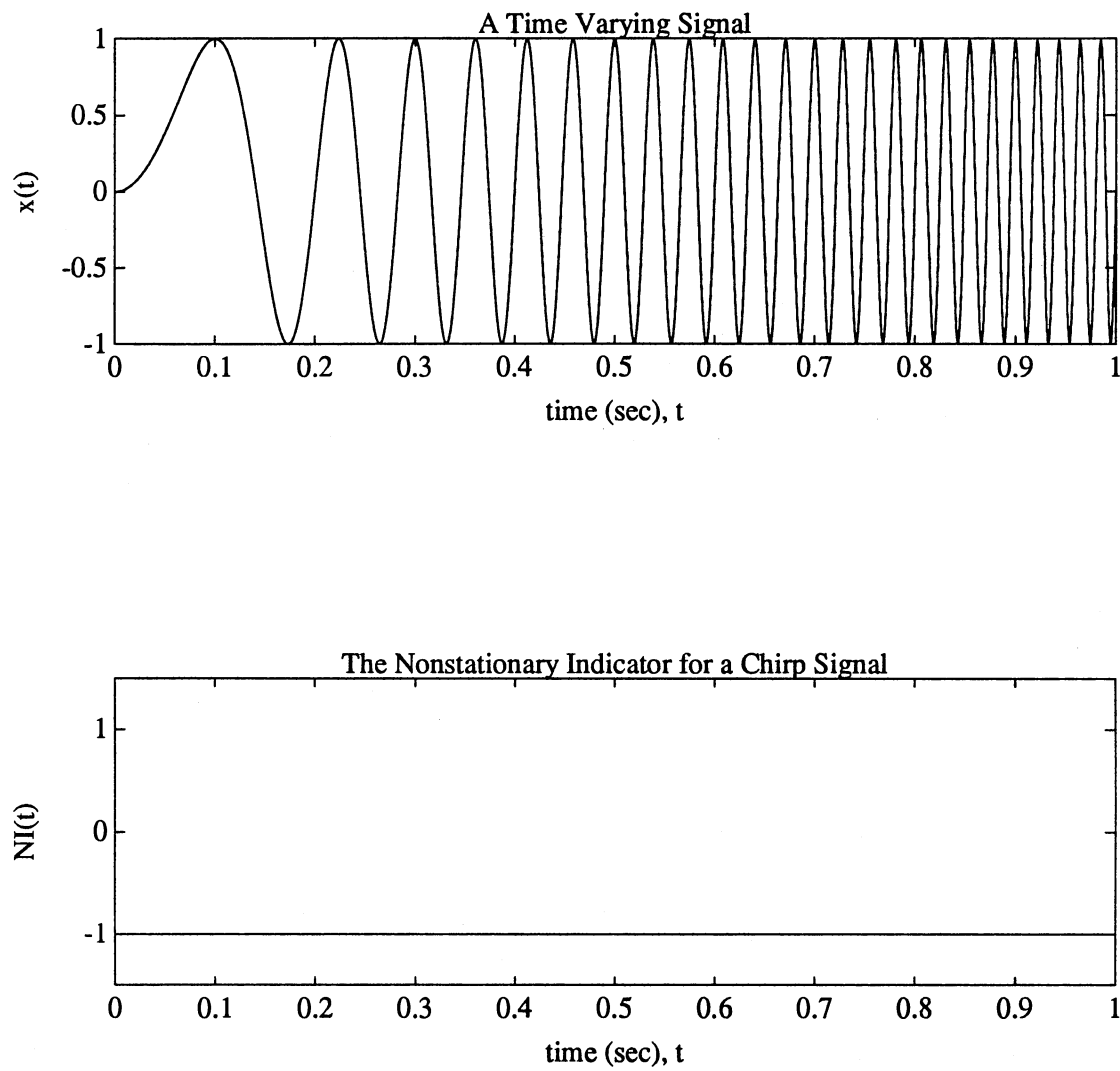


Figure V.5 The Ideal Nonstationary Indicator Performance for a Chirp Signal

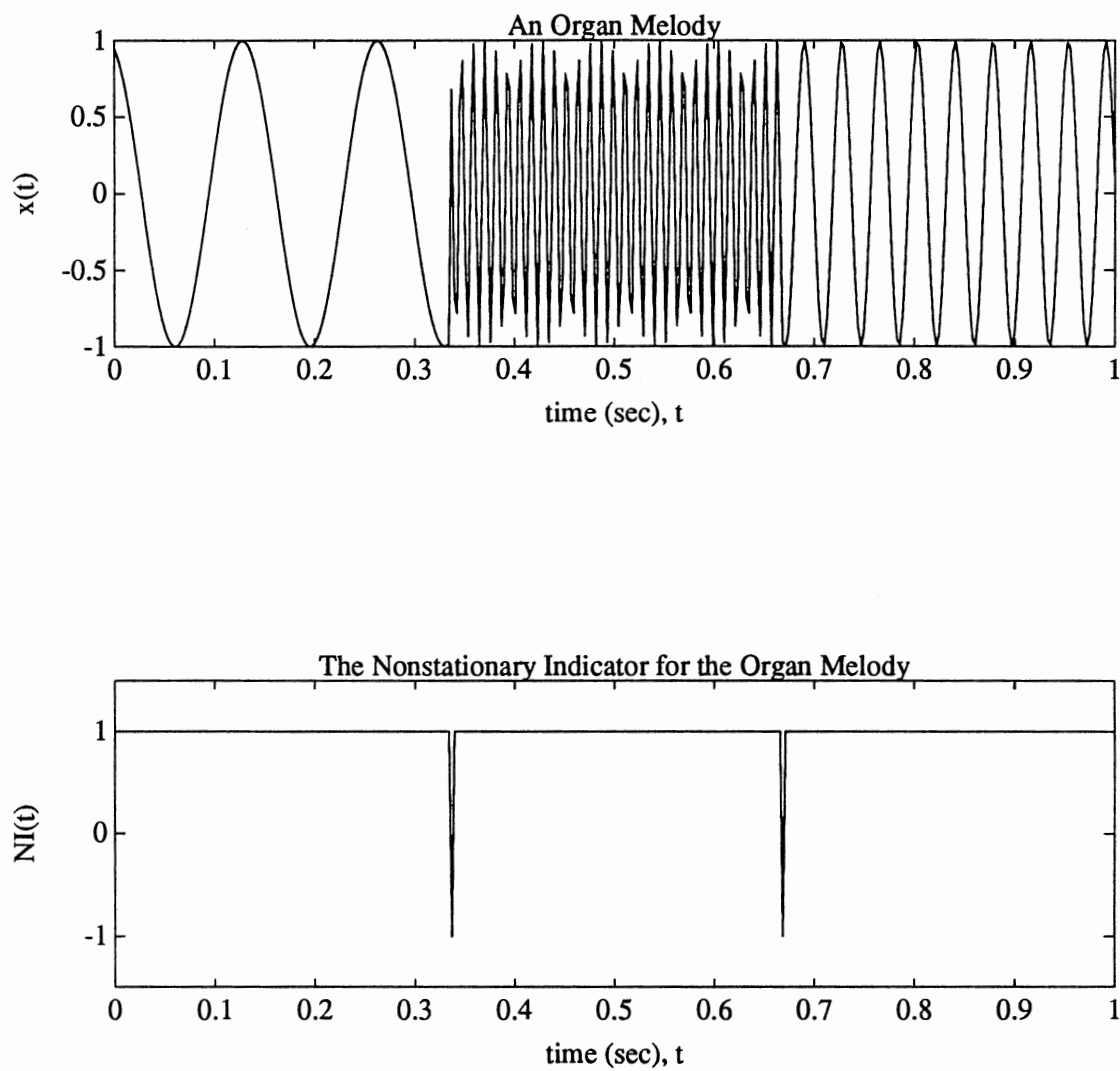


Figure V.6 The Ideal Nonstationary Indicator Performance for Three Concatenated Sinusoids

indicator equals -1. Thus, the example given in Figure V.6 illustrates how the nonstationary indicator applies to signal processing since the nonstationary indicator can tell when a signal changes.

Now that the ideal characteristics of the nonstationary indicator have been discussed, a method for computing the nonstationary indicator from Wigner distribution features is described. An abstract definition of the nonstationary indicator defines the nonstationary indicator as

$$NI(t) = g(f_1(IP_w(t)), f_2(IF_w(t)), f_3(VA_w(t)), f_4(SK_w(t)), f_5(KU_w(t))), \quad (V.23)$$

where g , f_1 , f_2 , f_3 , f_4 , and f_5 are arbitrary real functions. Clearly the $f_i(\cdot)$'s are functions of the Wigner distribution features. Although (V.23) is quite general, it does point out the basics of implementing the nonstationary indicator. First, the f_i functions act separately on each feature. These feature functions can describe a simple weighting of a feature or a complicated filtering operation performed on a feature. The main nonstationary indicator function, g , combines the feature function information into the actual nonstationary indicator output value. The function g will need to limit its output to $[-1, 1]$.

As an example of how the functions of (V.23) can be defined, one of the simplest nonstationary indicator implementations is described next. About the least complicated signal processing procedure is a linear combination of weighted inputs. A linear combination is applied to the nonstationary indicator by first defining the f_i functions with

$$f_i(x) = \alpha_i \cdot x + \beta_i, \quad (V.24)$$

where α_i and β_i are real numbers. Then the g function is defined by

$$g(x_1, x_2, x_3, x_4, x_5, \beta) = SGN(x_1 + x_2 + x_3 + x_4 + x_5 + \beta),$$

$$\text{where} \quad SGN(x) = \begin{cases} +1, & x \geq 0 \\ -1, & x < 0 \end{cases} \quad (V.25)$$

and β is a real number equalling the sum of all β_i . Thus, a simple nonstationary indicator implementation is

$$NI(t) = SGN(\alpha_1 IP_w(t) + \alpha_2 IF_w(t) + \alpha_3 VA_w(t) + \alpha_4 SK_w(t) + \alpha_5 KU_w(t) + \beta), \quad (V.26)$$

which is the signum of a weighted feature linear combination. This nonstationary indicator formulation is considered simple since the functions f_i and g are easily defined.

Notice that (V.25) constrains the nonstationary indicator's output to $[-1, 1]$ by including a nonlinear function. In addition, notice that if any of the Wigner distribution features do not contain useful information, then the appropriate α_i can equal zero.

The nonstationary indicator given in (V.26) is most likely too simple for real world applications. The reason for the simplicity can be seen by yet another example. Suppose only two Wigner distribution features, x_1 and x_2 , are used. Then (V.26) becomes

$$NI(t) = SGN(y(x_1, x_2)) = SGN(\alpha_1 x_1 + \alpha_2 x_2 + \beta). \quad (V.27)$$

The function inside the signum in (V.27) describes a line in the cartesian coordinate system. Thus, $y(x_1, x_2)$ can be displayed for a fixed α_1 , α_2 , and β . Setting $\alpha_1 = 1$, $\alpha_2 = 1$, and $\beta = -1$, Figure V.7 illustrates the line's graph. If $y(x_1, x_2) = x_1 + x_2 - 1$ is substituted in (V.27), then

$$NI(t) = SGN(x_1 + x_2 - 1). \quad (V.28)$$

Thus, if $x_1 = 1$ and $x_2 = 1$, then $NI(t) = 1$. Similarly, if $x_1 = -1$ and $x_2 = -1$, then

$NI(t) = -1$. By substituting in all possible x_1 and x_2 values into (V.28), the nonstationary indicator performs the mapping illustrated in Figure V.8 where the shaded area represents $NI(t) = -1$ and the blank area represents $NI(t) = 1$. Thus, the nonstationary indicator equals 1 to the right of the line $y(x_1, x_2)$ while the nonstationary indicator equals -1 to the left of the line $y(x_1, x_2)$. Due to the linear behavior involved with this nonstationary indicator, (V.27) is referred to as a linear separator.

The nonstationary indicator simple example shows that if the nonstationary indicator output can be separated by a line, then the nonstationary indicator implementation

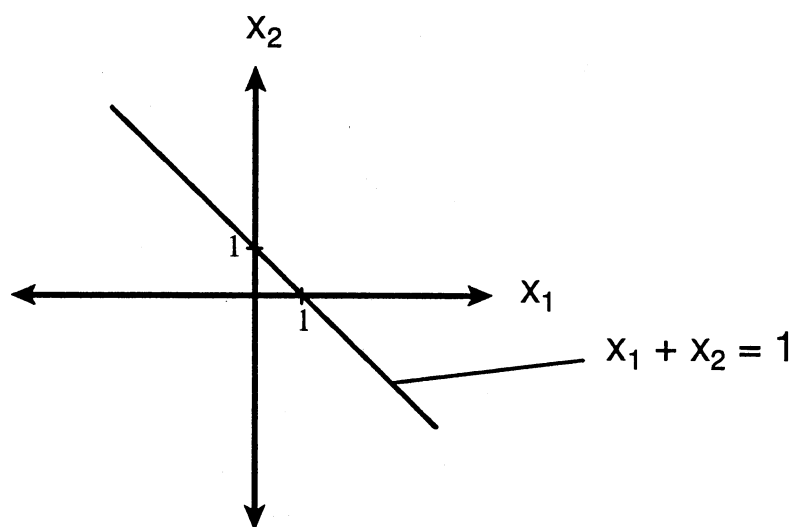


Figure V.7 The Line Used in a Simple Nonstationary Indicator Implementation

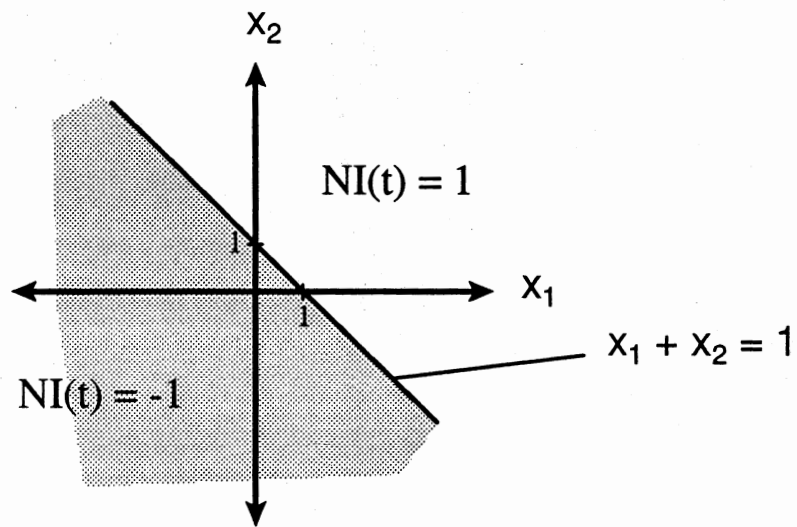


Figure V.8 Nonstationary Indicator Values for the Simple Example

given in (V.27) produces correct results for the appropriate α_1 , α_2 , and β . Similarly, if the nonstationary indicator can be separated by a hyperplane, then the nonstationary indicator implementation in (V.26) produces correct results for the appropriate x_1 , x_2 , x_3 , x_4 , x_5 , and β . Two problems exist, however. First, how are the values of α_i and β determined? The next section presents a technique, called training, for obtaining the α_i and β . Training requires a group of training signals and the desired associated nonstationary indicator values. Using the signals and the nonstationary indicator, α_i and β are found. Once the parameters are identified, (V.26) can then identify untrained signals. This procedure functions well if the nonstationary indicator output is linearly separable. However, what if the nonstationary indicator output is not linearly separable?

If the linear weighted combination nonstationary indicator, (V.26), does not perform correctly, then other functions for f_i and g must be developed. Although many potential formulations exist, the functions must be defined so that all parameters are computable. Thus, the functions f_i and g must be carefully chosen. In addition, the function f_i may change depending on the kind of output the nonstationary indicator produces. For instance, the simple example constrained the nonstationary indicator's output to either 1 or -1. Other implementations may require the nonstationary indicator to produce intermediate values between 1 or -1 so signals in-between stationary and nonstationary are recognizable.

For this thesis, the function g is implemented using a neural network and specifically, the multilayer perceptron. The multilayer perceptron was chosen because it is capable of arbitrarily mapping the five features to the nonstationary indicator output. Because the multilayer perceptron performs arbitrary mappings, features that do not provide useful nonstationary indicator information will not effect the output signal. Thus, the multilayer perceptron is significantly more powerful than the linear combination model which only maps linearly. In addition, the parameters contained within the multi-

layer perceptron are computed using the training procedure described for finding the linear coefficients. However, unlike the case where the linear coefficients can be obtained, the multilayer perceptron, using the backpropagation algorithm, may not implement the optimal system. The reason for performing suboptimally is due to the nonlinear functions that comprise the multilayer perceptron. However, the nonlinear functions enable the multilayer perceptron to implement any mapping. Thus, arbitrary mapping is traded for suboptimal performance. The reader should note, however, that even with a suboptimal performing method, the multilayer perceptron implements many systems that have never been implemented before. The next section describes the multilayer perceptron and backpropagation learning algorithm in detail since their operation forms the basis of the nonstationary indicator.

The Multilayer Perceptron

The Weighted Linear Combination

Since the multilayer perceptron is a complicated mapping technique, the multilayer perceptron discussion begins with the previously defined weighted linear combiner. By adding functions to the weighted linear combination, the multilayer perceptron is developed. Reviewing the multilayer perceptron by this approach allows for more intuition concerning the multilayer perceptron and also gives an idea concerning the difficulties found when using the multilayer perceptron. Recall that a weighted linear combination can be written as

$$y = \sum_{i=1}^N x_i w_i, \quad (\text{V.29})$$

where x_i represents the input, w_i represents the weights, and y , of course, is the output. In signal processing, (V.29) is called a finite impulse response (FIR) filter and is described

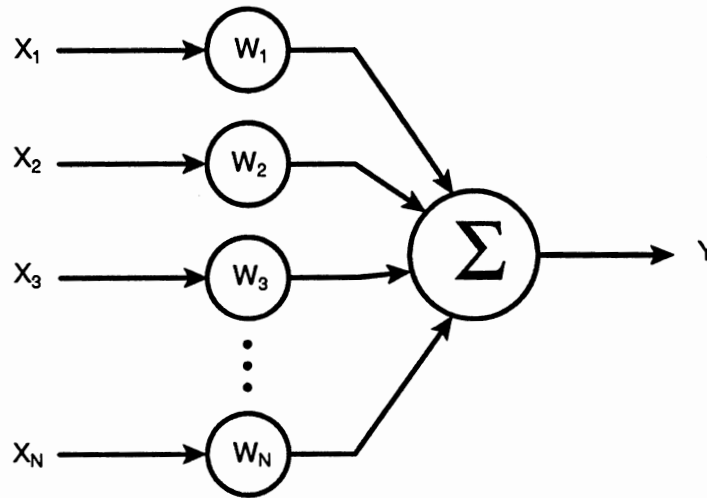


Figure V.9 A Linear Combination Network

pictorially in Figure V.9. The system described by (V.29) is general enough to produce many different algorithms. However, signal processing uses (V.29) for two main techniques: filtering and pattern recognition.

The filtering application arises from determining w_i that remove unwanted signals from x_i . The most common filter is probably the lowpass filter which can be defined for N weights by $w_i = 1/N$. Of course, many different filtering techniques can be implemented by defining different w_i . The pattern recognition application comes from using (V.29) in conjunction with the signum function given in (V.25). An example of a pattern recognition application is given in the previous section by (V.27).

Although many techniques exist for finding the values of w_i that produce different filters, finding values of w_i for a pattern recognition system is another story. Fortunately, a technique exists that produces w_i for either a filter or a pattern recognizer. This technique, called training, uses input values and desired output values to iterate a set of w_i that minimizes the squared error between the actual output and the desired output. The algorithm that implements the training process is called the least mean square (LMS) algorithm. Widrow and Hoff [Wid60] developed the LMS algorithm in 1962 to train a pattern recognition system. However, they quickly realized the LMS algorithm's potential in regular filtering [Wid85].

The LMS algorithm is given by

$$w_i^{(k+1)} = w_i^{(k)} + 2\mu\epsilon_i^{(k)}x_i^{(k)}$$

where

$$\epsilon_i^{(k)} = y_i^{(k)} - d_i^{(k)}. \quad (\text{V.30})$$

The superscript k represents the iteration number, ϵ_i represents a gradient estimate, μ is a convergence factor bounded in $[0, 2]$, and d_i is the desired output. The LMS algorithm is shown in Figure V.10 where the adaptive algorithm block represents (V.30) and the arrows piercing the weight circles imply the weights change over time. The LMS algorithm works by initially setting w_i to random values. Then the first input and desired output pair $(x_i^{(1)}, d_i^{(1)})$ is used with the current weights, $w_i^{(1)}$, to produce $y_i^{(1)}$ and $\epsilon_i^{(1)}$. Using $w_i^{(1)}$, $\epsilon_i^{(1)}$, and $x_i^{(1)}$, the new weights, $w_i^{(2)}$, are computed so that the squared error $(\epsilon_i^{(k)})^2$ is eventually minimized. The process repeats until the training process produces w_i that do not change or when no new $(x_i^{(i)}, d_i^{(i)})$ pairs exist. When the training process finishes, the weights, w_i , are then placed into the weighted linear combiner shown in Figure V.9 and given by (V.29).

The LMS algorithm is useful because the optimal set of weights, w_i , can always be found if enough data is made available. However, the word optimal is used here as

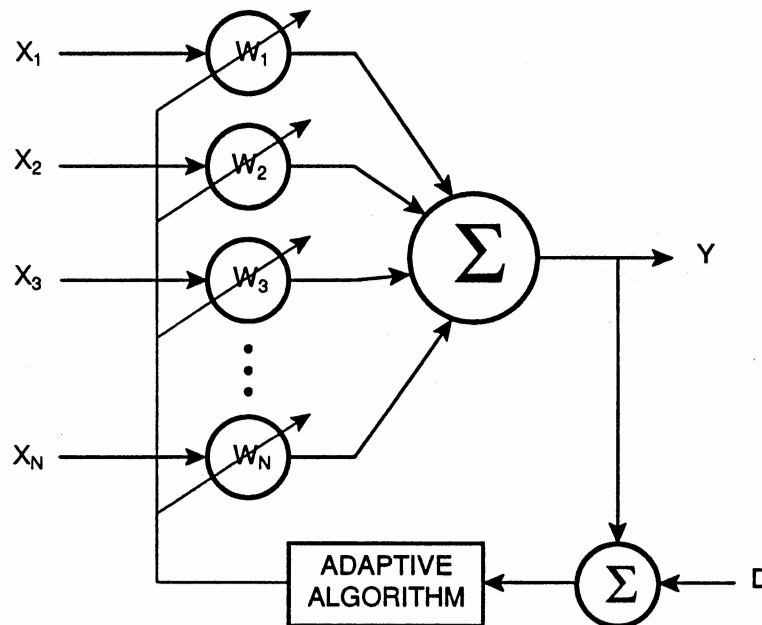


Figure V.10 An Adaptive Linear Combination Network

meaning that the minimum error can always be reached. Reaching the minimal error producing weights which does not say anything about how high the minimum error will be for a system. The reason for optimality is that the error surface, described by ϵ_i as a function of w_i , is convex and the LMS algorithm implements a gradient descent algorithm. Thus, as the number of iterations increases, the gradient descent homes in on the minimal squared error. What makes minimization work is the linear structure of the weighted linear combination (V.29). However, the linear structure of (V.29) is also quite limited, as was seen in the previous section when it was shown that only linearly separable pattern recognition systems could be produced. Thus, the weighted linear

combination along with the LMS algorithm is not capable of arbitrary mappings [Lip87]. How then can arbitrary mappings occur using the weighted linear combination's general structure and the LMS algorithm? Although many methods exist for extending the weighted linear combination, the Volterra filter and the multilayer perceptron will be discussed in this thesis.

The Volterra Filter

The Volterra filter is based on Volterra's expansion of input-output relationships given by

$$y = \sum_i w_i x_i + \sum_i \sum_j w_{ij} x_i x_j + \sum_i \sum_j \sum_k w_{ijk} x_i x_j x_k + \dots, \quad (\text{V.31})$$

where the number of summation groups represents the order number of the Volterra expansion. As an example, a first order Volterra expansion equals the weighted linear combiner (V.29). The second order Volterra expansion is given by

$$y = \sum_{i=1}^N w_i x_i + \sum_{i=1}^N \sum_{j=1}^i w_{ij} x_i x_j. \quad (\text{V.32})$$

Although the Volterra expansion has been utilized in nonlinear system theory [Scz80], the Volterra expansion has not been popular in signal processing implementations. The Volterra expansion's absence in signal processing is probably due to the extreme computational requirements of (V.31). However, these heavy computational requirements can be reduced by truncating the Volterra expansion to contain at most second order nonlinearities. The question is whether truncating the Volterra expansion still allows any advantages over the weighted linear combiner. The truncated Volterra expansion's advantage over the weighted linear combiner is best seen by example.

Recall that the weighted linear combination nonstationary indicator implementation given by (V.27) was $NI(t) = SGN(y(x_1, x_2))$ where $y(x_1, x_2) = \alpha_1 x_1 + \alpha_2 x_2 + \beta$. Converting the function y into a Volterra expansion is given by

$$V(x_1, x_2) = \alpha_1 x_1 + \alpha_2 x_2 + \beta + \alpha_{11} x_1 x_1 + \alpha_{12} x_1 x_2 + \alpha_{22} x_2 x_2, \quad (\text{V.33})$$

where three additional coefficients α_{11} , α_{12} , and α_{22} are introduced. If $\alpha_{11} = 1$, $\alpha_{12} = 1$, and $\alpha_{22} = 1$ along with the values of α_1 , α_2 , and β used as before, then

$$V(x_1, x_2) = x_1 + x_2 + x_1^2 + x_2^2 + x_1 x_2 = 1$$

so that
$$NI(t) = \text{SGN}(x_1 + x_2 + x_1^2 + x_2^2 + x_1 x_2 - 1). \quad (\text{V.34})$$

Thus, if $x_1 = 1$ and $x_2 = 1$, $NI(t) = \text{SGN}(4) = 1$. Similarly, if $x_1 = -1$ and $x_2 = -1$ then

$NI(t) = \text{SGN}(-2) = -1$. These two results are the same as for the weighted linear combiner. However, when inputting more values of x_1 and x_2 into (V.34), the mapping changes drastically from a linear separable case to the nonlinear mapping given in Figure V.11. As with Figure V.8, the shaded area of Figure V.11 indicates $NI(t) = -1$ while the blank area represents $NI(t) = 1$. Therefore, Figure V.11 indicates that the second order Volterra expansion of the simple nonstationary indicator example produces a very complicated decision surface. Because the Volterra expansion creates such a complicated decision surface in \mathfrak{R}^2 , the Volterra expansion allows very complicated decision surfaces in hyperspace, \mathfrak{R}^N .

Not only does the second order Volterra expansion produce complicated decision surfaces but the LMS algorithm can find the Volterra expansion's weight. The reason that the LMS algorithm applies to the nonlinear Volterra expansion is that the second order Volterra expansion is just the linear combination of x_1 , x_2 , x_{11} , x_{12} , x_{22} , and β instead of just x_1 , x_2 , and β . Thus, the nonlinear components of the Volterra expansion are treated as additional inputs. The only problem with the Volterra expansion is that an infinite order is required for arbitrary mappings [Scz80]. Since a second order Volterra expansion is far from an infinite order, the second order Volterra expansion may not be capable of implementing a nonstationary indicator. However, because the mapping between the Wigner distribution features and the nonstationary indicator will probably be very

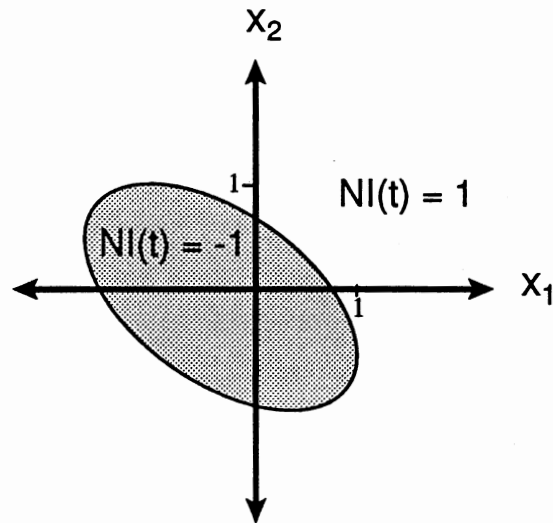


Figure V.11 Nonstationary Indicator Values for the Volterra Expansion Simple Example

nonlinear, the second order Volterra expansion will not be directly used. Instead, the second order Volterra expansion idea will be applied to the multilayer perceptron which is described next.

The Nonlinear Combination

Developing the multilayer perceptron begins not by adding inputs as was done with the Volterra expansion but by including a nonlinear function to the basic weighted linear combination. The weighted nonlinear combination is given by

$$y_i = f(o_i)$$

where

$$o_i = \sum_{j=1}^N x_j w_j + \beta_i, \quad (V.35)$$

$f(o_i)$ is some nonlinear function, and β is a bias term. Equation (V.35) generalizes the weighted linear combination since $f(o_i) = o_i$ produces the weighted linear combination. However, (V.35) allows much more complicated systems since f can be practically anything. In fact, if $f(o_i) = \text{SGN}(o_i)$, then the simple nonstationary indicator example can be implemented directly. However, because certain types of responses are desired from (V.35), the function f is normally restricted to 1:1 continuous functions. The 1:1 restriction is a necessary condition placed by the method which updates the nonlinear combination weights (presented in the next paragraph). The continuity restriction allows many mathematical manipulations.

The questions that remain are how the added nonlinear functions effect training and how they aid arbitrary mapping. First, training can no longer be implemented by the LMS algorithm given in (V.30) but must be implemented by

$$w_i^{(k+1)} = w_i^{(k)} + 2\mu\epsilon^{(k)}f'(x_i^{(k)}w_i^{(k)})$$

$$\text{or} \quad w_i^{(k+1)} = w_i^{(k)} + 2\mu\epsilon^{(k)}f'(o^{(k)}), \quad (\text{V.36})$$

where $f'(o^{(k)})$ represents the derivative of the nonlinear function f with respect to the input $o^{(k)}$ [Pao89]. Figure V.12 illustrates how the basic weighted linear combination system of Figure V.10 is modified to include the nonlinear function (note that $o(n)$ in the Figure V.12 represents $o^{(k)}$). The reader should note that the derivative in (V.36) requires f continuous. In addition, for (IV.36) to yield optimal solutions, f must be a convex function since the composition of convex functions is itself convex. Thus, restricting the nonlinear function f to be 1:1 and continuous allows the nonlinear LMS algorithm (V.36) to implement the training process.

Currently, two functions typically implement the nonlinear function f : the sigmoid function and the hyperbolic tangent function [Kli91]. The sigmoid function is given by

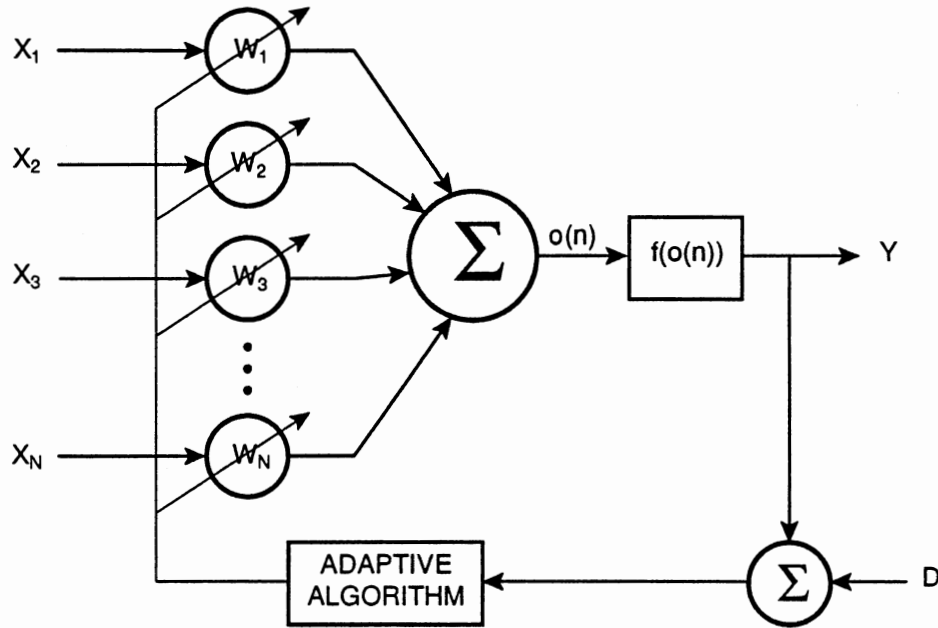


Figure V.12 An Adaptive Nonlinear Combination Network

$$f_1(x) = \frac{1}{1 - e^{-x}}, \quad (\text{V.37})$$

while the hyperbolic tangent function is given by

$$f_2(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (\text{V.38})$$

These two functions implement the nonlinear function of the weighted nonlinear combination because both functions closely approximate the sigum function, and both function derivatives are computable using only the function's output value. Figure V.13

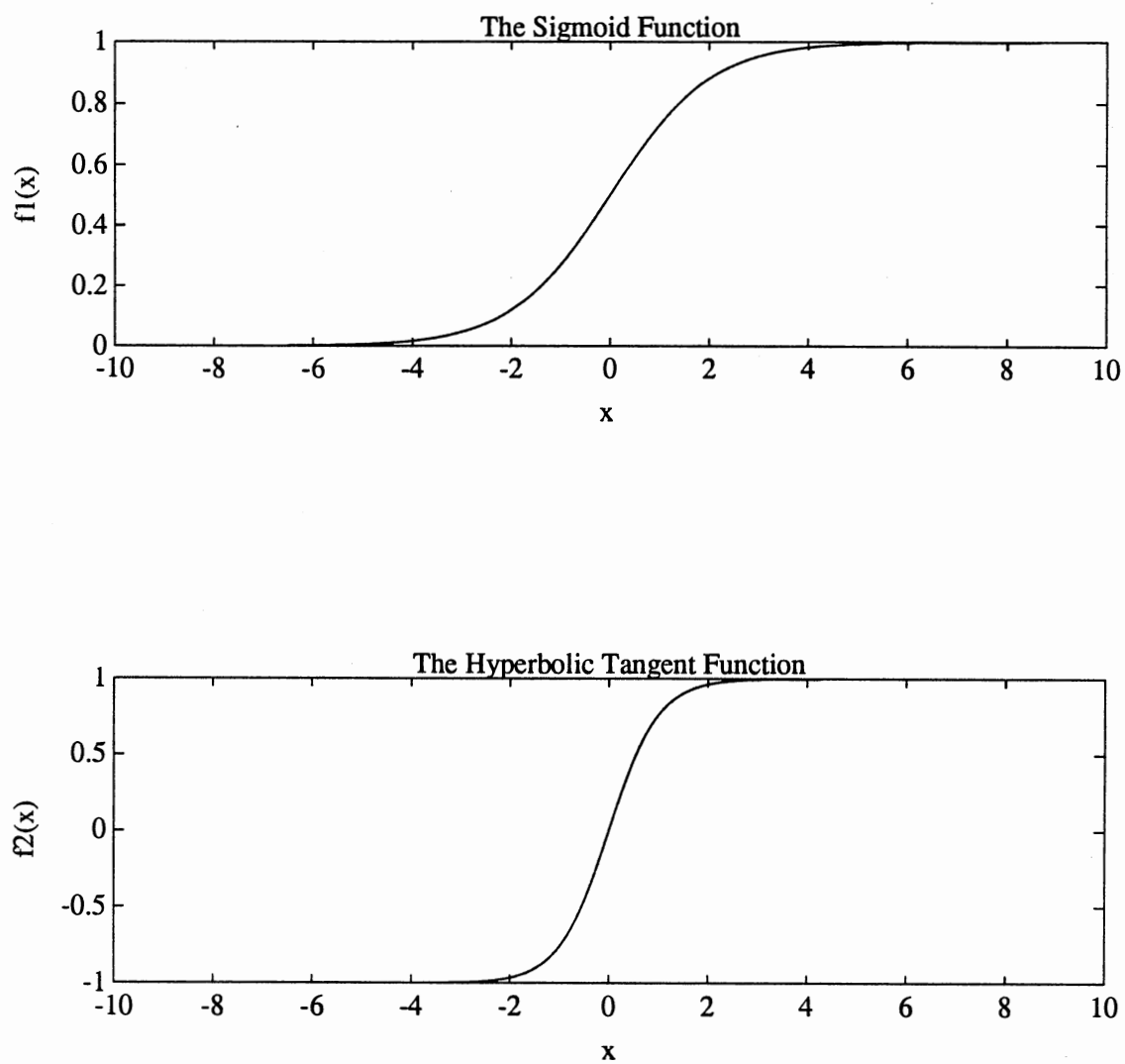


Figure V.13 Graphs of the Sigmoid Nonlinear Function and the Hyperbolic Tangent Nonlinear Function

illustrates the graph of the sigmoid function and the hyperbolic tangent function for inputs ranging from -10 to 10 . Notice how both functions behave similar to the signum function with the hyperbolic tangent function being a better approximation. The two functions are different, however, because

$$\lim_{x \rightarrow -\infty} f_1(x) = 0 \text{ and } \lim_{x \rightarrow \infty} f_1(x) = 1,$$

while

$$\lim_{x \rightarrow -\infty} f_2(x) = -1 \text{ and } \lim_{x \rightarrow \infty} f_2(x) = 1. \quad (\text{V.39})$$

As previously stated, one of the advantages of these two functions is that their derivatives can be written in terms of the function's output. The sigmoid derivative is given by

$$f_1'(x) = f_1(x)(1 - f_1(x)), \quad (\text{V.40})$$

while the hyperbolic tangent derivative is computed by

$$f_2'(x) = 1 - f_2^2(x). \quad (\text{V.41})$$

Substituting these derivative formulae into the nonlinear LMS algorithm (V.36) produces the sigmoid LMS algorithm given by

$$w_i^{(k+1)} = w_i^{(k)} + 2\mu\epsilon^{(k)}x^{(k)}f(o^{(k)})(1 - f(o^{(k)})) \quad (\text{V.42})$$

and the hyperbolic tangent LMS algorithm given by

$$w_i^{(k+1)} = w_i^{(k)} + 2\mu\epsilon^{(k)}x^{(k)}(1 - f^2(o^{(k)})). \quad (\text{V.43})$$

Most implementations seem to favor the sigmoid function which is found frequently in neural models [Rum81]. However, the hyperbolic tangent function has been found to be a better choice in some instances; thus, both may be useful [Kli91].

Even though the weighted nonlinear combination extends the capabilities of the weighted linear combination, arbitrary mappings are still not possible. Arbitrary mappings are possible [Lip87] by using the weighted nonlinear combination in a network shown in Figure V.14. This network is called the multilayer perceptron. Each circle in Figure V.14 represents the summation operation and the nonlinear function for the weighted nonlinear combination. Each branch in Figure V.14 has an implicit weight

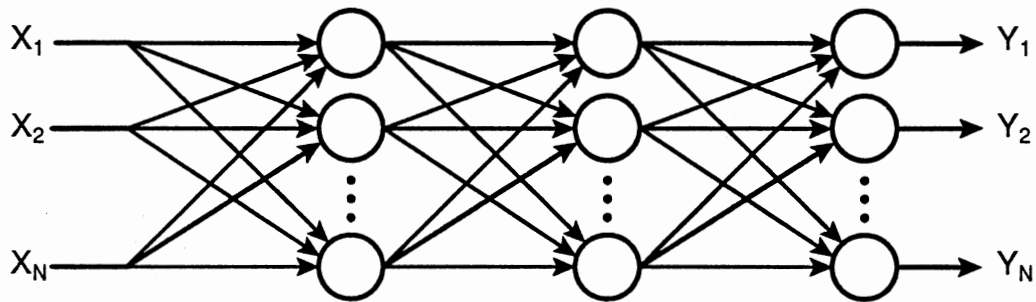


Figure V.14 A Multilayer Perceptron

associated with it. Thus, the multilayer perceptron, given in Figure V.14, contains three layers each containing weighted nonlinear combinations with multiple outputs. The term multilayer perceptron comes from the fact that each weighted nonlinear combination could be a perceptron (a weighted nonlinear combination with signum nonlinear functions) [Ros62]. The perceptron was quite popular in the 1960's, but no technique was found that could efficiently train the weights of the multilayer perceptron. Not until the early 1980's when Rumelhart developed the backpropagation algorithm did a method exist that indicated how multilayer perceptron weights are updated [Rum81].

The problem with updating multilayer perceptron weights has to do with the fact that the hidden layers (layers to the left of the outputs Y_i 's in Figure V.14) do not have access to the output errors. However, using several chain rules, equations for training all weights can be developed. First, the output layer is trained using the ordinary nonlinear LMS algorithm, which can be rewritten as

$$w_{lm}^{(k+1)} = w_{lm}^{(k)} + \mu x_l^{(k)} \delta_m^{(k)}$$

where

$$\delta_m^{(k)} = f'(o_m^{(k)}) \epsilon_m^{(k)}, \quad (\text{V.44})$$

m represents which output of the output layer, and l indicates which output layer input.

The new part of the algorithm is found for the hidden layers and is given by

$$w_{lm}^{(k+1)} = w_{lm}^{(k)} + \mu x_l^{(k)} \delta_m^{(k)}$$

where

$$\delta_m^{(k)} = f'(o_m^{(k)}) \sum_{i=1}^N \delta_i^{(k)} w_{il}^{(k)}. \quad (\text{V.45})$$

In (V.45), l represents the current hidden layer's input, m represents the current hidden layer's output, and i represents the $l + 1$ layer's input. Notice that the only difference between (V.44) and (V.45) is that $\sum_{i=1}^N \delta_i^{(k)} w_{il}^{(k)}$ replaces $\epsilon_m^{(k)}$ in (V.44). This replacement is due to the fact that hidden layers do not have direct access to $\epsilon_m^{(k)}$ but must obtain $\epsilon_m^{(k)}$ information by backpropagating $\delta_i^{(k)}$'s (each $\delta_i^{(k)}$ is either the actual output error, $\epsilon_i^{(k)}$, or the output error of a hidden layer).

The pair of backpropagation equations, (V.44) and (V.45), are generalizations of the LMS algorithm. In fact, some researchers refer to (V.44) and (V.45) as the generalized LMS algorithm. The main problem with the multilayer perceptron and the backpropagation algorithm is that the nonlinear behavior of the multilayer perceptron does not provide optimal solutions. The reason for suboptimal performance is that the multilayer perceptron is not a convex function. Thus, the backpropagation algorithm still performs a conjugate gradient type approach to minimize the least square error given only a suboptimal solution, i.e., the solution corresponding to a local minima. However,

experimentation has showed that the backpropagation algorithm does do a fair job of obtaining reasonable solutions. If the exact optimal solution is necessary, techniques exist for obtaining optimal solutions. First, the obvious method would experiment with new initial starting points and choose the final set of weights which produce the smallest square error. The other choice is based on a concept from material science call annealing. Simulated annealing perturbs the weights of the multilayer perceptron based on Boltzman's probability distribution [Pre86]. As such, the multilayer perceptron weights are adjusted frequently when learning begins but adjust less frequently as time goes on. It has been shown that simulated annealing can produce optimal solutions [Szu86].

The multilayer perceptron training procedure is different than the LMS training procedure, because passing all the data though one time may not provide a robust system. Instead, a group of (x_i, d_i) pairs, called the training set, is passed through the backpropagation algorithm several times. Each pass is called an epoch, and a typical training session may take several thousand epochs to achieve reasonable results. Thus, training begins by setting all weights to random numbers. Then the first epoch propagates $x_i^{(1)}$ through the multilayer perceptron. The error computed by $d_i^{(1)}$ and $y_i^{(1)}$ is then backpropagated to modify the weights. Either the weight change is made instantaneously or the weight change is averaged with the remaining epoch weight changes and applied at the end of the epoch. Regardless of the error update method, this process of propagating the input and back propagating the output error is continued until no new data exists. The entire process continues until the multilayer perceptron is performing adequately or until the multilayer perceptron fails to converge. If the backpropagation algorithm fails to converge, the entire process can be repeated using a different initial weight setting, or the number of network neurons can be changed.

Arbitrary Mappings

The next question which requires an answer is if the multilayer perceptron can perform arbitrary mappings. Evaluating the multilayer perceptron's learning capabilities is an ongoing research topic that just recently has been partially answered [Pao89]. One partial solution is based on Hilbert's 13th problem [Lor76]. This problem arose in 1900 when David Hilbert, a famous mathematician, delivered a lecture before the International Congress of Mathematicians. Hilbert discussed 23 problems of mathematical research which researchers should attempt to solve in the new century. In Hilbert's 13th problem, he conjectured that given any continuous function of three variables, the function cannot be obtained by finitely many substitutions of continuous functions of two variables.

Kolmogorov, in 1958, rejected Hilbert's claim using a general theorem [Kol57]. Kolmogorov proved that there exists fixed continuous functions $\phi_{pq}(x)$ on $I \in [0, 1]$ so that each continuous function f on I^n can be written in the form

$$f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} g_q \left(\sum_{p=1}^n \phi_{pq}(x_p) \right), \quad (\text{V.46})$$

where $g_q(\cdot)$ are properly chosen continuous functions of one variable.

So far the link between (V.46) and the multilayer perceptron is not evident; but in 1965, Sprecher [Spr65] improved Kolmogorov's theorem by replacing the function of two variables with a function of one variable and a coefficient in the following theorem. Sprecher's theorem is presented next.

Theorem V.1: There exists $(2n + 1)$ continuously increasing functions $\phi_1, \phi_2, \dots, \phi_{2n+1}$ mapping $[0, 1] \rightarrow \Re$ and having the additional property that for any continuous function f of n real variables, there exists a continuous function, $g : [0, n] \rightarrow \Re$ with real coefficients λ_p such that

$$f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} g_q \left(\sum_{p=1}^n \lambda_p \phi_q(x_p) \right)$$

for all values of $x_1, x_2, \dots, x_n \in [0, 1]$. †††

As an example of this theorem, let $f(x_1, x_2) = x_1 x_2$. If $g_q(\xi) = \frac{1}{5} e^{\xi}$, $\phi_i(x_i) = \ln(x_i)$, and $\lambda_i = 1$,

then

$$\begin{aligned} f(x_1, x_2) &= \sum_{q=1}^5 g_q \left(\sum_{p=1}^2 \lambda_p \phi_p(x_p) \right) = \sum_{q=1}^5 \frac{1}{5} e^{\sum_{p=1}^2 \ln(x_p)} \\ &= e^{\ln(x_1) + \ln(x_2)} = e^{\ln(x_1 x_2)} = x_1 x_2. \end{aligned} \quad (\text{V.47})$$

Thus, Theorem V.1 allows writing a nonlinear function, $x_1 x_2$, in terms of the summation of function compositions (the summation corresponds to the summing of exponential functions and the composition arises from the natural logarithm functions). Notice that much of the theorem's flexibility was not necessary for this example. Thus, Theorem V.1 is quite powerful.

This form of Hilbert's problem was shown by Hecht-Nielsen to be almost identical to the multilayer perceptron if g_q is taken to be the sigmoid or hyperbolic tangent function, ϕ_q equals one, and λ_p represents the weights [Hec89]. However, Kolmogorov's theorem and Sprecher's improvement state that the g functions only exists and may not be a sigmoid or hyperbolic tangent function. Nonetheless, Theorem V.1 shows the potential of a two layer multilayer perceptron, even though the theorem does not prove arbitrary mappings. However, other techniques which use the Stone-Weierstrass Theorem and allowing an infinite number of hidden units, have proved that a two layer multilayer perceptron can perform arbitrary mappings [Cot90].

The purpose of delving into a multilayer mapping theorem was to show that although any number of layers can be used, only one hidden layer is necessary for arbitrary mappings. Therefore, the multilayer perceptron illustrated in Figure V.15 will be used to implement the nonstationary indicator since given enough hidden elements this multilayer perceptron performs arbitrary mappings. Using one hidden layer compared to

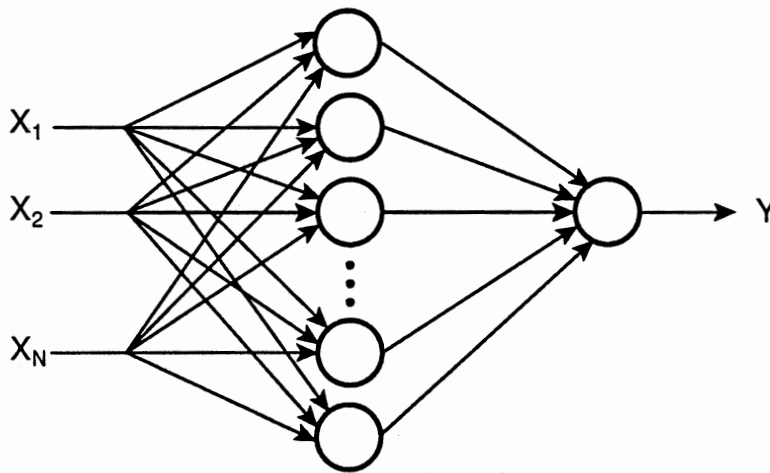


Figure V.15 A Multilayer Perceptron with One Hidden Layer and One Output

several hidden layers drastically reduces the training time. Additionally, the multilayer perceptron that implements the nonstationary indicator executes faster. The disadvantage of using only one hidden layer is that many neurons (weighted nonlinear combinations) might be necessary. To reduce the potential number of hidden neurons, the Volterra expansion is applied to the hidden layer.

The Volterra Expansion Multilayer Perceptron

The Volterra expansion multilayer perceptron modifies the normal two layer multilayer perceptron by computing the Volterra expansion of the hidden layer output. The additional hidden layer outputs then are input into the output layer. For a normal multilayer perceptron, if $f(o_i), \dots, f(o_N)$ represent the N outputs of the hidden layer for a

two layer multilayer perceptron, then they also represent the input to the multilayer perceptron's output layer. As such, the output, y , for the output layer would then be computed by

$$y = f\left(\sum_{i=1}^N f(o_i)w_i + \beta\right). \quad (\text{V.48})$$

The Volterra expansion modifies (V.48) using the following output equation

$$y = f\left(\sum_{i=1}^N f(o_i)w_i + \sum_{i=1}^N \sum_{j=1}^N f(o_i)f(o_j)w_{ij} + \beta\right), \quad (\text{V.49})$$

where w_{ij} does not necessarily equal w_{ji} . The additional $\sum_{i=1}^N \sum_{j=1}^N f(o_i)f(o_j)w_{ij}$ term is, of course, from the Volterra expansion. Hopefully, adding the Volterra expansion term into the hidden layer allows the two layer multilayer perceptron to admit arbitrary mappings using relatively few hidden neurons.

The only problem with including the second order Volterra expansion into the hidden layer output is training the weights before and after the hidden layer. The weights residing before the hidden layer utilize the hidden layer outputs to compute the backpropagated error. The additional Volterra expansion hidden layer outputs are included in the δ calculations just like (V.44) and (V.45) indicate. Thus, the Volterra expansion hidden layer outputs do not cause too much trouble for updating weights occurring before the hidden layer. The only problem remaining is how the Volterra expansion hidden layer output weights are updated.

The Volterra expansion hidden layer output weight update procedure is derived by looking at (V.44) and by noticing the only terms not known are the derivatives of the Volterra expansion hidden layer outputs. Using the fact that the ordinary hidden layer output derivatives can be written as

$$\frac{dg(o_i)}{dx_i} = \frac{df(o_i)}{dx_i}, \quad (\text{V.50})$$

where $g(x_i)$ represents the hidden layer output, the Volterra expansion hidden layer output derivative can be written as

$$\frac{dg(o_i, o_j)}{dx_i} = \frac{d}{dx_i}(f(o_i)f(o_j)). \quad (\text{V.51})$$

Note that $f(o_j)$ is a function of x_j but not of x_i . Equation (V.51) is divided into two disjoint groups where $i \neq j$ comprises one group and $i = j$ makes up the other. Then

$$\begin{aligned} \frac{dg(o_i, o_j)}{dx_i} &= (f'(o_i)f(o_j)) + (0 \cdot f(o_i)) \\ &= f'(o_i)f(o_j) \end{aligned} \quad (\text{V.52})$$

when $i \neq j$. Alternately, if $i = j$, then

$$\begin{aligned} \frac{dg(o_i, o_j)}{dx_i} &= \frac{df^2(o_i)}{dx_i} = 2f'(o_i)f(o_i) \\ &= 2f'(o_i)f(o_j). \end{aligned} \quad (\text{V.53})$$

Thus,

$$\frac{dg(o_i, o_j)}{dx_i} = \begin{cases} f'(o_i)f(o_j), & \forall i \neq j \\ 2f'(o_i)f(o_j), & \forall i = j \end{cases}. \quad (\text{V.54})$$

The Volterra expansion hidden layer output derivatives are almost identical to normal hidden layer output derivatives except for a scaling factor and the multiplication by $f(o_j)$. To compare the Volterra expansion hidden layer outputs to normal hidden layer outputs, the following shows a comparison. For a sigmoid nonlinearity,

$$f'(\cdot) = \begin{cases} f(o_i)(1-f(o_i)) & , HL \\ \begin{cases} f(o_i)(1-f(o_i))f(o_j), & \forall i \neq j \\ 2f(o_i)(1-f(o_i))f(o_j), & \forall i = j \end{cases} & , VEHL \end{cases}, \quad (\text{V.55})$$

where $f'(\cdot)$ represents the general derivative function in (V.44), *HL* represents the multilayer perceptron hidden layer, and *VEHL* represents the Volterra expansion multilayer perceptron. Similarly, the hyperbolic tangent nonlinearity produces

$$f'(\cdot) = \begin{cases} 1 - f^2(o_i) & , HL \\ \begin{cases} (1 - f^2(o_i))f(o_j), & \forall i \neq j \\ 2(1 - f^2(o_i))f(o_j), & \forall i = j \end{cases} & , VEHL \end{cases} \quad (V.56)$$

Thus, the Volterra expansion hidden layer's weights can now be updated.

Now that the two layer multilayer perceptron and the two layer Volterra expansion multilayer perceptron have been discussed, arbitrarily mapping the Wigner distribution features to the nonstationary indicator may be possible. However, before implementing the nonstationary indicator, the software designed for the two layer multilayer perceptron and the two layer Volterra expansion multilayer perceptron should be verified. In addition to validating the multilayer perceptron, the multilayer perceptron's ability to map nonlinear functions can also be tested.

Multilayer Perceptron and Volterra Expansion Multilayer Perceptron Testing

The multilayer perceptron and the Volterra expansion multilayer perceptron software was tested to learn the following nonlinear function

$$\begin{aligned} d = f(x, y, z) &= |x + y + z| - \lfloor |x + y + z| \rfloor \\ &= \text{FRAC}(x + y + z) \quad , \end{aligned} \quad (V.57)$$

where x , y , and z are real numbers, $\lfloor \cdot \rfloor$ represents the floor operation, and $\text{FRAC}(\cdot)$ represents the fractional part of a real number. Thus, the nonlinear function adds three real numbers and chops off the whole number and the number's sign. Although summing three numbers together is linear, the fractional operation is nonlinear since it computes an absolute value and also truncates some values.

The example training set is developed by setting x , y , and z to uniform distributed random numbers which contain zero mean and are bounded between -1 and 1 . Because (V.57) represents the desired signal, d is computed using the random inputs. Thus, the $(x^{(i)}, y^{(i)}, z^{(i)}, d^{(i)})$ comprise the training sample set. Before applying the training samples

to the neural network software, note that because the fractional part of any real number cannot exceed or equal 1. Thus, because $[0, 1]$ bounds the sigmoid function values, the sigmoid nonlinearity is used in the multilayer perceptron.

The first test covers the Volterra expansion multilayer perceptron. The Volterra expansion multilayer perceptron configuration is setup so that the hidden layer contains 16 neurons. Because (V.57) contains 3 inputs and 1 output, this neural network is referred to as a 3-16-1 Volterra expansion multilayer perceptron. With this configuration, 217 weights will require training when taking into account the bias weights. A good rule of thumb for determining the necessary number of training samples that allows good neural network generalization is [Kli91]

$$\#training\ sets > 5(\#weights). \quad (V.58)$$

If (V.58) is not adhered to, then the network trains itself to map each individual input instead of discovering general characteristics between the input and the output. As such, 2000 training samples should be more than sufficient using 217 weights.

The 2000 training samples were fed into the Volterra expansion multilayer perceptron software along with parameter settings of $\mu = .3$ and $\alpha = .7$. Recall that μ represents the convergence factor whose values must be between 0 and 2. Letting $\mu = .3$ forces the network to train conservatively since μ determines how much of the gradient estimate to trust at each iteration. The other parameter, α , is called the momentum term. Although momentum is not included in (V.42), momentum just adds $\alpha(w^{(i)} - w^{(i-1)})$ to $w^{(i+1)}$ in (V.42). The momentum term also forces conservative behavior from the network because momentum causes the gradient descent to continue along the same path.

Figure V.16 illustrates the 3-16-1 Volterra expansion multilayer perceptron training results. The graph shows how the average error between the desired output and the actual output decreases as the number of epochs increases. According to the top graph in Figure V.16, the 3-16-1 Volterra expansion multilayer perceptron is on its way towards convergence. In fact, at epoch number 2000, the average error is only 0.021. To check whether

the 3-16-1 Volterra expansion multilayer perceptron generalizes the nonlinear function, a different training set consisting of 10000 samples was propagated through the network with the weights set. The average error for untrained data was 0.032. The untrained data average error is typically greater than the trained data average error but not more than one order of magnitude greater. Therefore, the 3-16-1 Volterra expansion multilayer perceptron seems to have learned the nonlinear function (V.57).

The other major piece of software implements the multilayer perceptron without the Volterra expansion included. Using the same configuration as the Volterra expansion multilayer perceptron training session (3-16-1), using the same training data, and not changing the parameters produces the average error graph located in the middle of Figure V.16. Notice that the average error is quite high, even after 2000 epochs. Thus, the 3-16-1 multilayer perceptron used here is not capable of learning the nonlinear function, (V.57). However, notice that the multilayer perceptron contains fewer weights than a Volterra expansion multilayer perceptron. In fact, for a 3-16-1 Volterra expansion multilayer perceptron only 81 weights are used instead of the 3-16-1 Volterra expansion multilayer perceptron's 217 weights. Therefore, to give the regular multilayer perceptron the same capacity for learning (V.57), a multilayer perceptron containing 221 weights is tested.

A 3-44-1 multilayer perceptron, which contains 221 weights, was trained using the same training data and the identical parameter settings as the 3-16-1 Volterra expansion multilayer perceptron. The bottom graph in Figure V.16 shows the average network error as a function the epoch number. Notice that the 3-44-1 multilayer perceptron seems to be learning but at a much slower rate than the 3-16-1 Volterra expansion multilayer perceptron. This result is fairly consistent for different training sets. However, the Volterra expansion multilayer perceptron performs no better than a comparable multilayer

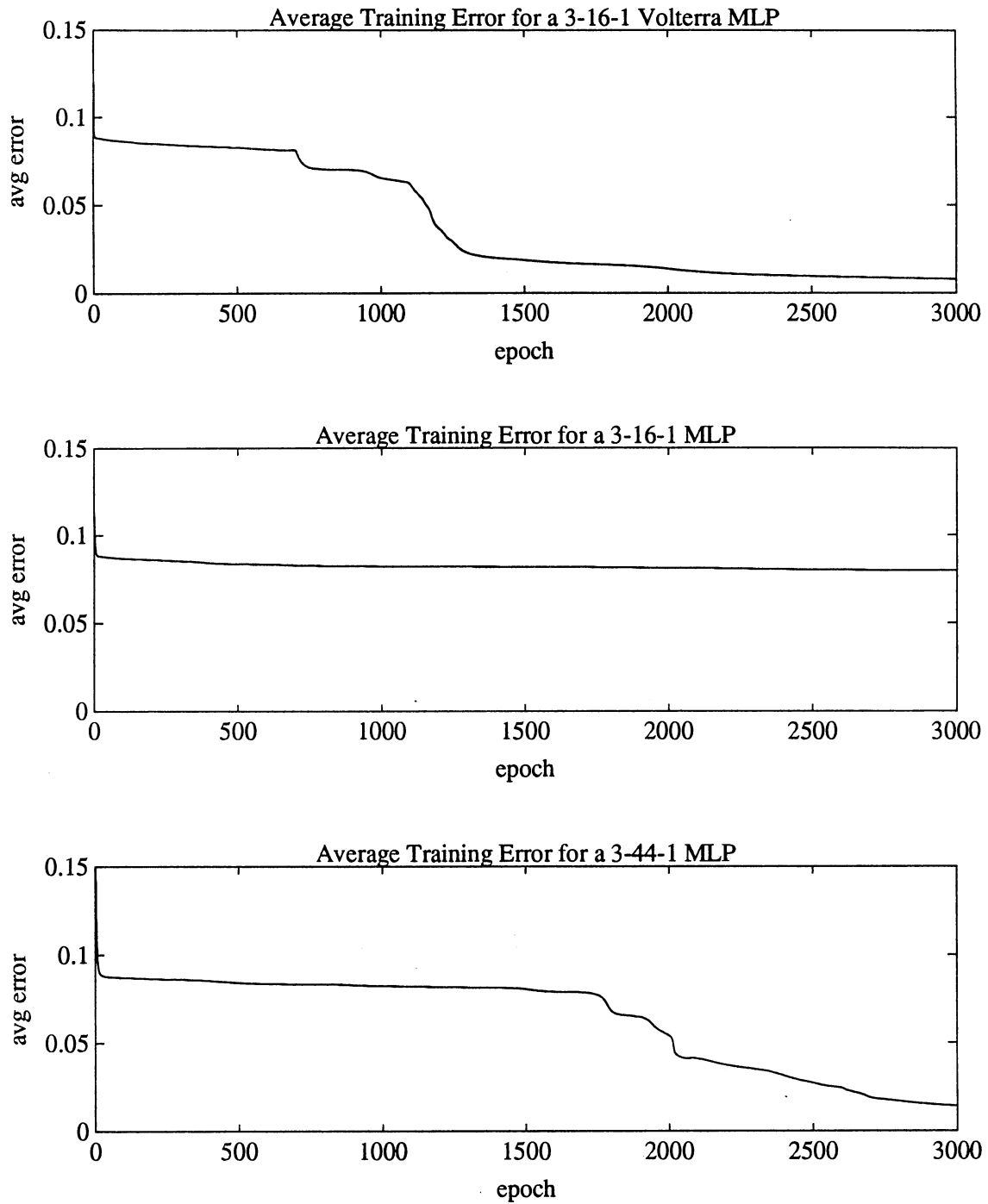


Figure V.16 The Nonlinear Function Average Training Error for a 3-16-1 Volterra Expansion Multilayer Perceptron, a 3-16-1 Multilayer Perceptron, and a 3-44-1 Multilayer Perceptron.

perceptron on some other examples. Thus, the Volterra expansion multilayer perceptron may or may not perform better than a multilayer perceptron for the nonstationary indicator.

The testing example indicates that all software performs as expected since both the Volterra expansion multilayer perceptron and the multilayer perceptron learned the nonlinear function (V.57). This example also brings up a few other interesting points. First, although the Volterra expansion multilayer perceptron trained faster than the multilayer perceptron, no firm conclusion can be made in general. Thus, both the Volterra expansion multilayer perceptron and the multilayer perceptron will attempt to implement the nonstationary indicator. The second point involves the graphs found in Figure V.16. Notice the 3-16-1 Volterra expansion multilayer perceptron error and the 3-44-1 multilayer perceptron error indicate that several hundred epochs can pass before significant decreases of error occur. Thus, both the Volterra expansion multilayer perceptron and the multilayer perceptron must train for several thousand epochs. Next, the neural network is used to map Wigner distribution features to the nonstationary indicator.

A Nonstationary Indicator Implementation

Now that a technique exists which arbitrarily maps inputs to outputs, implementing the nonstationary indicator is possible. Because the Wigner distribution features may contain time varying spectral energy information, using Wigner distribution features as the input to a multilayer perceptron makes sense. Thus, the nonstationary indicator could be implemented simply by generating a set of input samples $X(k) = (IP_w(k), IF_w(k), VA_w(k), SK_w(k), KU_w(k))$ along with a set of desired output samples $NI(k)$ and then training a multilayer perceptron. If any of these five features fails to provide time varying spectral energy information, then the neural network is capable of "turning off" the appropriate features so that the failed features do not contribute to the nonstationary indicator output. Early investigation determined that this simple setup was not possible

because no relative temporal information is available to the multilayer perceptron. The whole idea behind the nonstationary indicator is to state when the signal's characteristics change over time. As such, the input to the nonstationary indicator must contain information from different times so the multilayer perceptron can perform comparisons. Therefore, the input is changed to accommodate the multilayer perceptron better and to use the time varying spectral energy information more efficiently.

First, recall that in Chapter III the Wigner distribution features seem to include time varying spectral information. The time varying spectral information is in the form of changing values of the features. However, by just letting the multilayer perceptron's input equal $X(k)$, the neural network does not receive any feature change information. Thus, the multilayer perceptron input should be modified to somehow include more Wigner distribution feature information. Two main methods accomplish this goal. First, the Wigner distribution features can be processed to enhance feature changes. The obvious method of enhancing features is to pass the Wigner distribution features through a highpass filter or a differentiator. However, because differentiating a signal enhances noise, an FIR differentiator that attenuates high frequency values but still estimates the derivative would be used. The, differentiated signals with or without the original signals could then act as the multilayer perceptron's input. The only problem with this method is that the derivative may not contain time varying spectral energy information. What if the time varying spectral information is actually more global? For example, if $X(k)$ differs from $X(k + 3)$ in a particular way, then the derivative method would not include this information. Additionally, if the feature derivatives are useful, the neural network will learn this trait and set its weights to compute the derivatives. Therefore, the first method probably is not suitable for the nonstationary indicator implementation.

The second method forces the multilayer perceptron to learn if Wigner distribution feature changes contain time varying spectral information by inputting several training samples at one time. Inputting several training samples at one time, which I call a *meta-training sample*, is described as

$$X^{(N)}(k) = \begin{pmatrix} X\left(k - \left\lfloor \frac{N}{2} \right\rfloor\right) \\ X\left(k - \left\lfloor \frac{N}{2} \right\rfloor + 1\right) \\ \vdots \\ X\left(k + \left\lfloor \frac{N}{2} \right\rfloor - 1\right) \\ X\left(k + \left\lfloor \frac{N}{2} \right\rfloor\right) \end{pmatrix}, \quad (\text{V.59})$$

where N is an odd integer that represents the number of different training samples and $X^{(N)}(k)$ represents the meta-training sample at time k . Notice that the meta-training sample is centered so that $NI(k)$ is computed using past and future Wigner distribution feature values. The advantage of this method is that if any temporal relationships exist between the Wigner distribution features, the multilayer perceptron will find them. If the derivative method was used, then temporal relationships might not be found.

The only unresolved issue with (V.59) is how $NI(k)$ should be defined. Two choices are available for choosing the meta-training sample based $NI(k)$. First, the simple method sets $NI(k)$ to what it would have been if the input consisted only of $X(k)$. This method ignores the fact that anything different is going on with the input. Using this method is fine for signals which are either strictly stationary or nonstationary. However, since temporal information is presented to the multilayer perceptron, temporal information should also be indicated by the output. As an example, say $N = 3$, $X(k - 1)$ implies $NI(k - 1) = 1$, $X(k)$ implies $NI(k) = 1$, and $X(k + 1)$ implies $NI(k) = -1$. Thus, although the center training sample comes from a stationary signal and would ordinarily produce

$NI(k) = 1$, there is a nonstationary training sample also present at the input. Since a distinct difference exists between an input containing training samples all from a stationary source and an input containing a mixture of training samples, the nonstationary indicator output should indicate these differences. Thus, a reasonable method for setting $NI(k)$ in the context of a meta-training sample is by using the average $NI(k)$ value over the input times. Thus, for a meta-training sample $X^{(N)}(k)$,

$$NI^{(N)}(k) = \frac{1}{N} \sum_{i=k-\lfloor \frac{N}{2} \rfloor}^{k+\lfloor \frac{N}{2} \rfloor} NI(i), \quad (\text{V.60})$$

where $NI^{(N)}(k)$ indicates a meta-training sample based nonstationary indicator. Notice that if $N = 1$, then the meta-training sample collapses into a training sample while (V.60) reduces to an ordinary $NI(k)$. One of the advantages of (V.60) is that by monitoring $NI^{(N)}(k)$ transitions between stationary and nonstationary, signals show up as values between -1 and 1 . Thus, (V.60) may actually provide more information than just the ordinary nonstationary indicator.

The second issue that must be addressed concerning the multilayer perceptron's input is the variety of magnitudes that each feature contains. Because the features are based on spectral energy, feature values change depending on originating signal's power. For example, two speech signals that contain the same sound but differ in loudness would have very different Wigner distribution feature values even though both signals contain basically the same information. In order to help multilayer perceptron training, inputs typically are normalized so that differences such as loudness do not interfere with analyzing significant relationships between the input and the output. Thus, the Wigner distribution features are normalized. However, what normalizing technique should be used? Two basic methods for normalizing Wigner distribution features exist. For the first method, the Wigner distribution itself could be normalized by dividing each Wigner distribution coefficient by the maximum absolute Wigner distribution value. However,

this method forces the Wigner distribution to contain more global information. Wigner distribution global information corresponds to how the Wigner distribution changes over long periods of time. Because the nonstationary indicator should detect signal characteristic changes immediately, a more local normalizing technique is desired. Thus, the second method normalizes the Wigner distribution features instead of the Wigner distribution itself. In addition, not only are the Wigner distribution features normalized, but each feature is normalized separately. Because multiple input training samples will be used, each Wigner distribution feature is normalized with respect to the meta-training sample. Thus, each feature in $X^{(N)}(k)$ is divided by the maximum feature magnitudes of $X^{(N)}(k)$. Therefore, $X^{(N)}(k) \in [-1, 1]$. Because each Wigner distribution feature is normalized, the multilayer perceptron should relate the Wigner distribution features to the nonstationary indicator easier.

One additional problem remains concerning the inputs to the multilayer perceptron. Recall that all of the Wigner distribution features, except for $IP_w(t)$, contain a division which may not be defined. As such, when the divisor is near zero, the particular feature value becomes artificially large. Because the Wigner distribution features contain positive and negative values, many near zero values exist. Because normalizing very large values causes the other values to approach zero, the erroneous very large feature values should be removed. Thus, before normalizing the Wigner distribution features, the features are median filtered. A median filter removes spurious values that are nowhere near the mean value. Specifically, the instantaneous frequency and the variance are processed using a three element median filter. The skewness and the kurtosis, which contain two possible near zero divisors, are filtered using a five element median filter. Thus, even though the median filter removes some Wigner distribution feature information, something has to remove the spurious values or the outliers.

Now that the nonstationary indicator implementation has been discussed, the proposed nonstationary indicator can be tested. Testing begins with simple examples and works up to more difficult examples if the nonstationary indicator shows potential. The simple example signal consists of 25 sinusoids and 25 chirps signals that have been randomly concatenated together. Figure V.17 shows the first half of the signal along with the first half of the ideal nonstationary indicator. Only half of the each signal is shown in Figure V.17 because plotting the entire signal on one graph hid the underlying structure of the example signals. Similarly, Figure V.18 shows the second half of both the test signal and the ideal nonstationary indicator. Figures V.17 and V.18 are provided simply to show the data used to train the nonstationary indicators. Notice that the nonstationary indicator graph does not indicate the use of the meta-training sample based nonstationary indicator. However, because the number of input training samples has not yet been determined, the nonstationary indicator averaging occurs in the neural network software. The signals displayed in Figure V.17 and V.18 attempted to train a 5-20-1 multilayer perceptron and a 5-16-1 Volterra expansion multilayer perceptron using $\mu = .3$, $\alpha = .7$, and $N = 7$. These parameter values were chosen from approximately 100 different variations because they seemed to provide the lowest training error and the shortest training time. In no way are these parameters optimal, but they should allow for good training of the nonstationary indicator.

The average training error for the 5-20-1 multilayer perceptron, up to epoch number 300, is shown in Figure V.19. The corresponding average training error for the 5-16-1 Volterra expansion multilayer perceptron is also provided in Figure V.19. Only 300 average errors are graphed to simplify comparing the training rates of the multilayer perceptron and the multilayer perceptron. However, both nonstationary indicators did eventually converge asymptotically. Notice that the multilayer perceptron and the Volterra expansion multilayer perceptron train at approximately the same rate. Thus, for this particular example, adding the Volterra expansion to the multilayer perceptron does

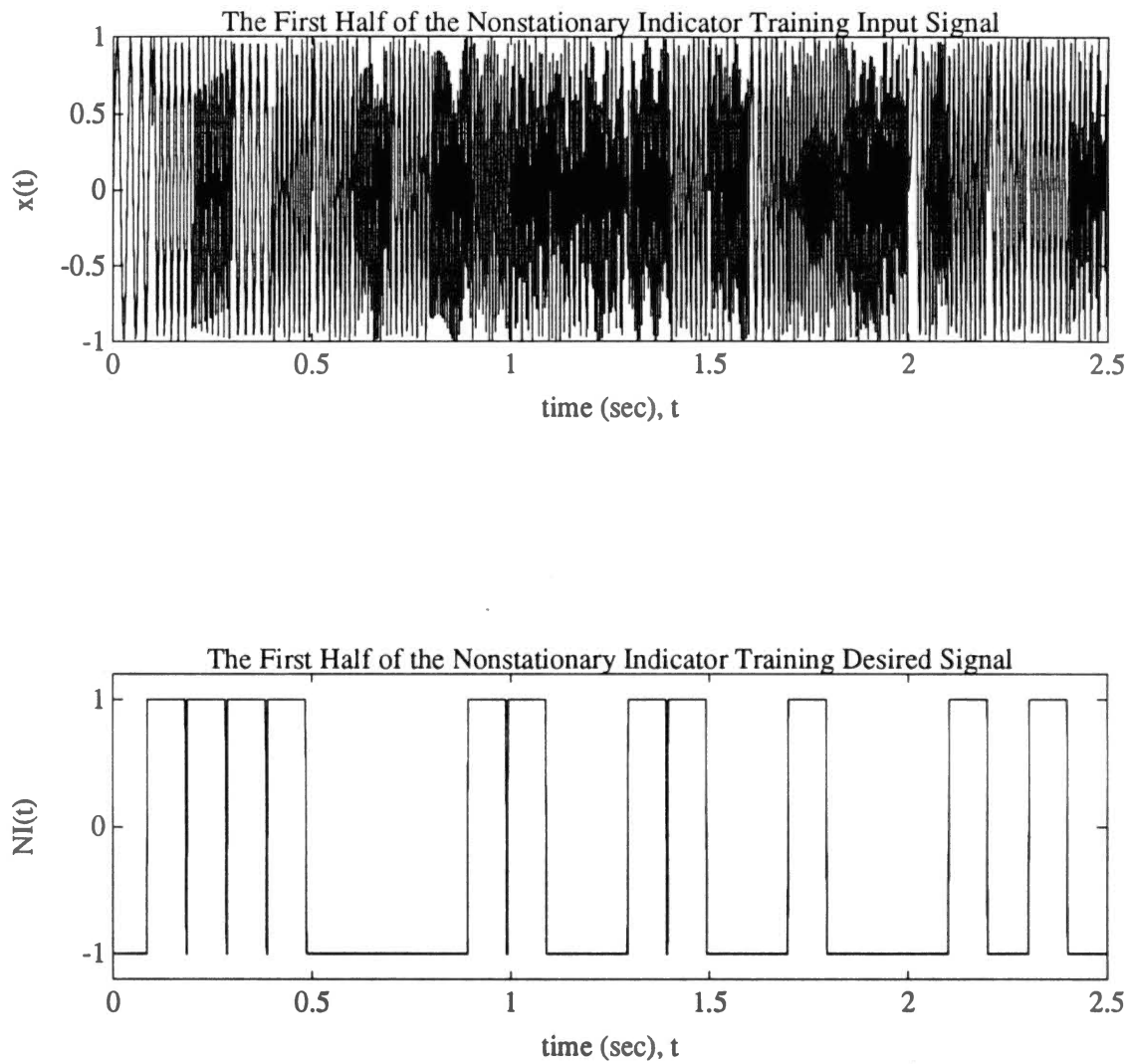


Figure V.17 A Test Signal Consisting of Sinusoids and Chirp Signals and the Ideal Nonstationary Indicator

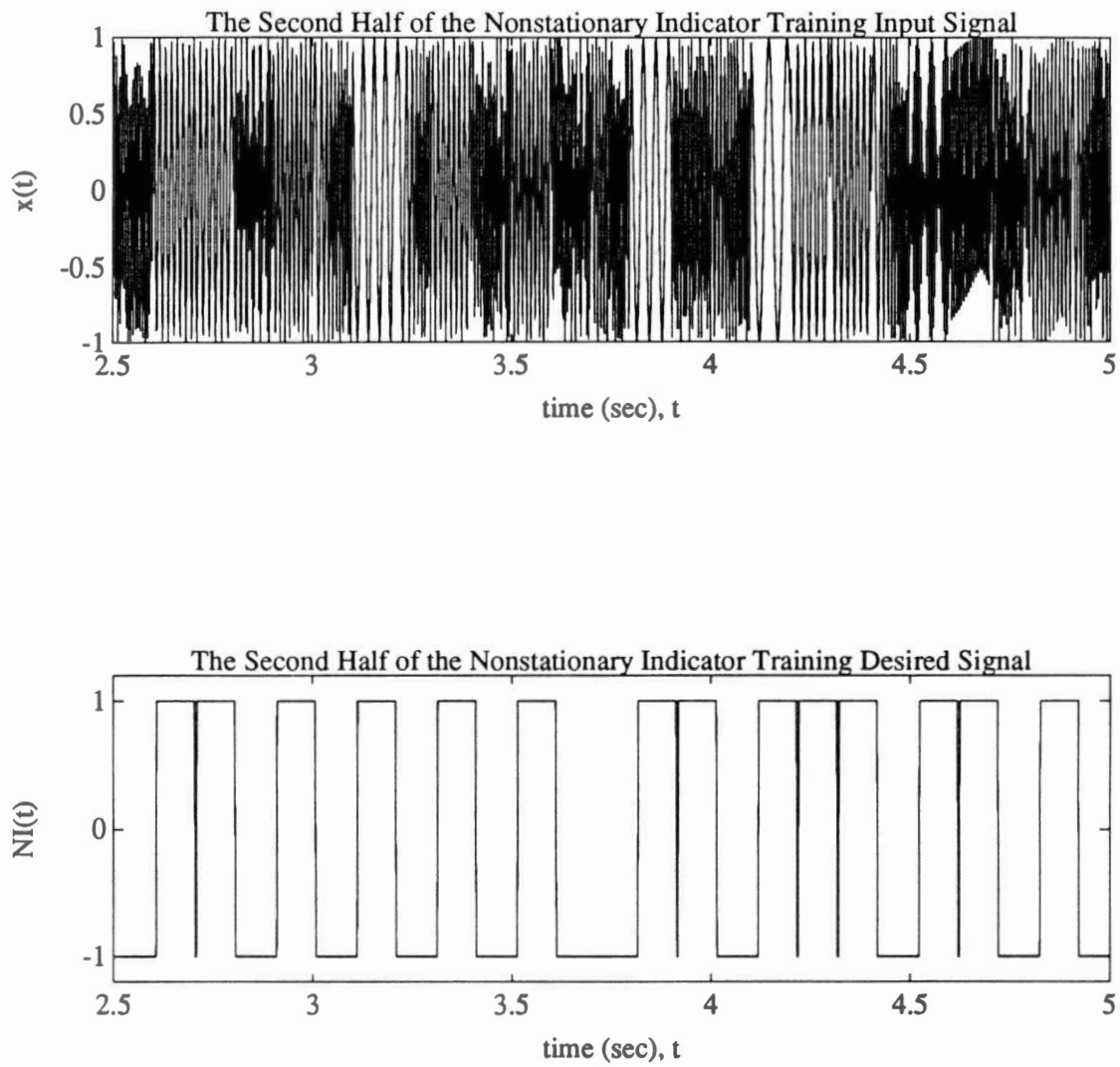


Figure V.18 A Test Signal Consisting of Sinusoids and Chirp Signals and the Ideal Nonstationary Indicator

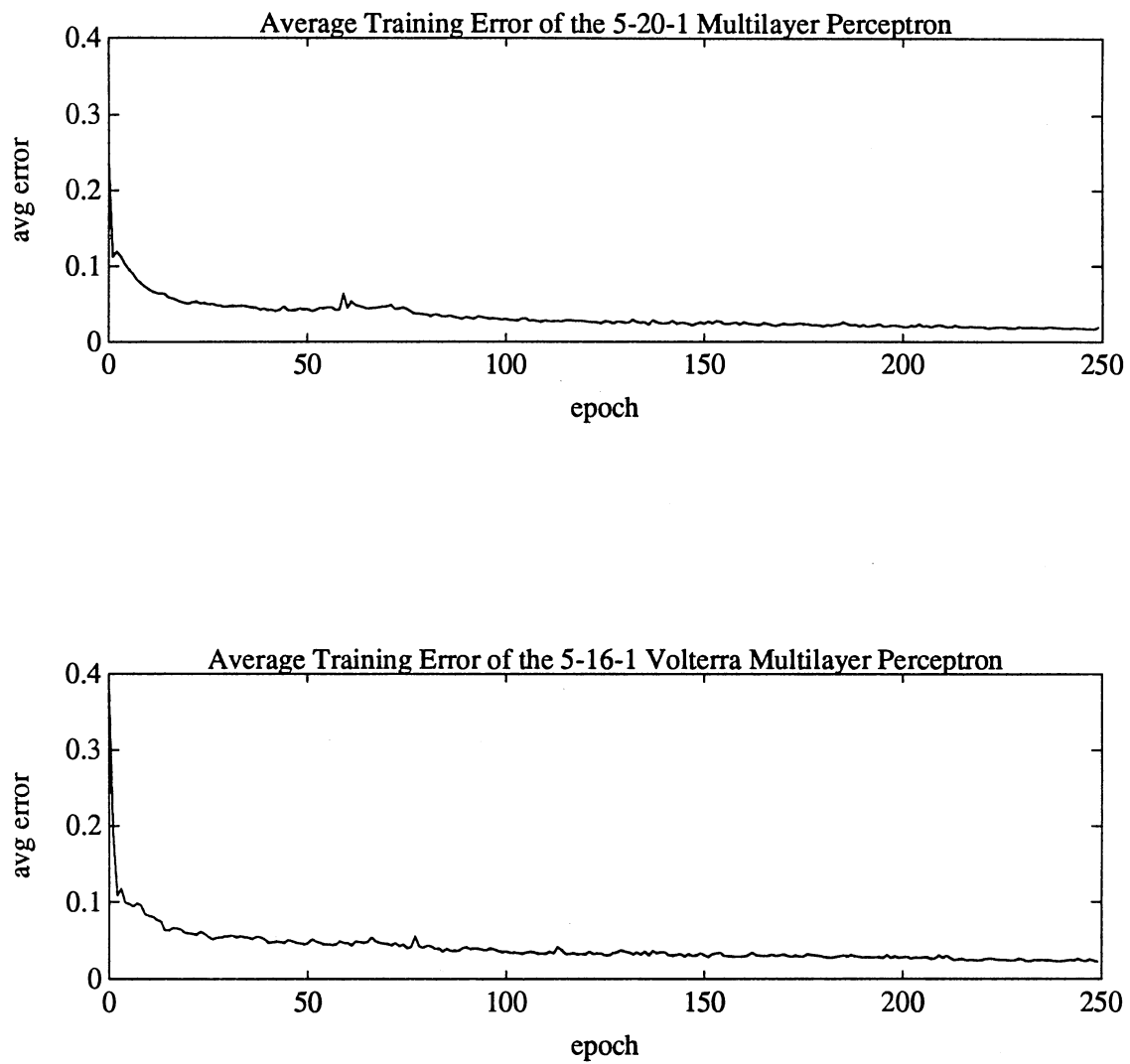


Figure V.19 Average Training Errors for a 5-20-1 Multilayer Perceptron and a 5-16-1 Volterra Expansion Multilayer Perceptron

not seem to speed-up training time. Although the final errors appear small (in fact both converged), the only way to check whether the multilayer perceptron and the Volterra expansion multilayer perceptron generalize the time varying spectral energy information is to investigate the nonstationary indicator's output.

Figure V.20 illustrates the multilayer perceptron, the Volterra expansion multilayer perceptron, and the ideal nonstationary indicator output to the first half of the trained meta-training samples. Additionally, Figure V.21 illustrates the second half of these signals. Recall that both nonstationary indicator implementations should perform almost perfectly on the trained data. However, the multilayer perceptron and the Volterra expansion multilayer perceptron outputs are exceptionally noisy. Although the basic output structure resembles the ideal nonstationary indicator, the large amount of noise (especially for nonstationary portions of the signal) indicates that both implementations had difficulty in generalizing the nonstationary indicator. Thus, the input may not correspond well to the output. This conclusion is substantiated by investigating how the multilayer perceptron and Volterra expansion multilayer perceptron nonstationary indicators perform on untrained data.

Figure V.22 provides the multilayer perceptron, the Volterra expansion multilayer perceptron, and the ideal nonstationary indicator output to the first half of the untrained meta-training samples. Additionally, Figure V.23 gives the second half of these signals. If the multilayer perceptron and Volterra expansion multilayer perceptron generalized the basic structure between the Wigner distribution features and the nonstationary indicator output, then the outputs displayed in Figures V.22 and V.23 should not vary significantly from the ideal output. Unfortunately, the outputs are extremely noisy for nonstationary portions of the signal and relatively noisy for the stationary portions. In fact, comparing Figures V.22 and V.23 to Figures V.20 and V.21 shows that the untrained data outputs are significantly worse than the trained data outputs. This also indicates that the multilayer perceptron and the Volterra expansion multilayer perceptron were not able to

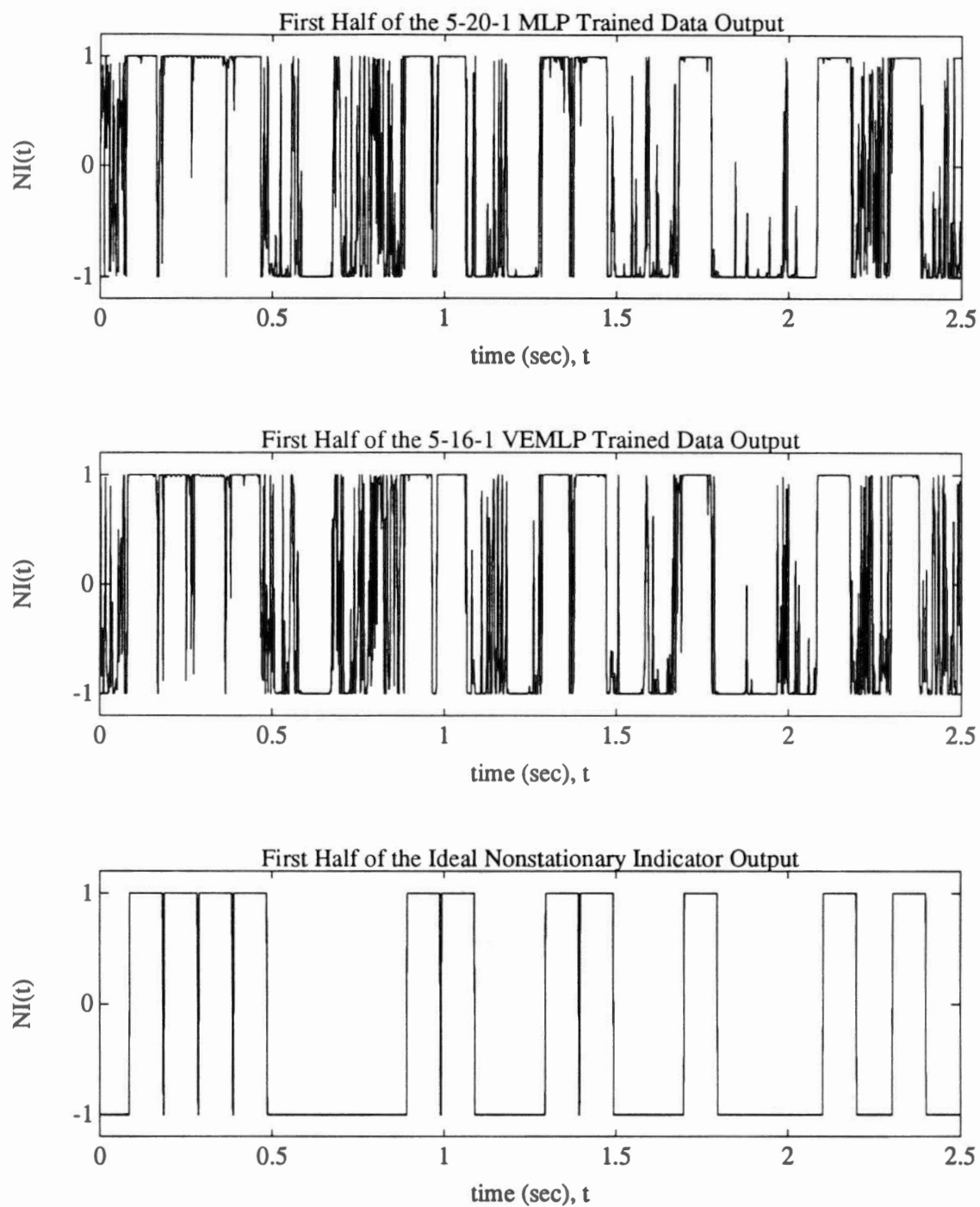


Figure V.20 The Nonstationary Indicator Output of a 5-20-1 Multilayer Perceptron and a 5-16-1 Volterra Expansion Multilayer Perceptron on the Trained Data.

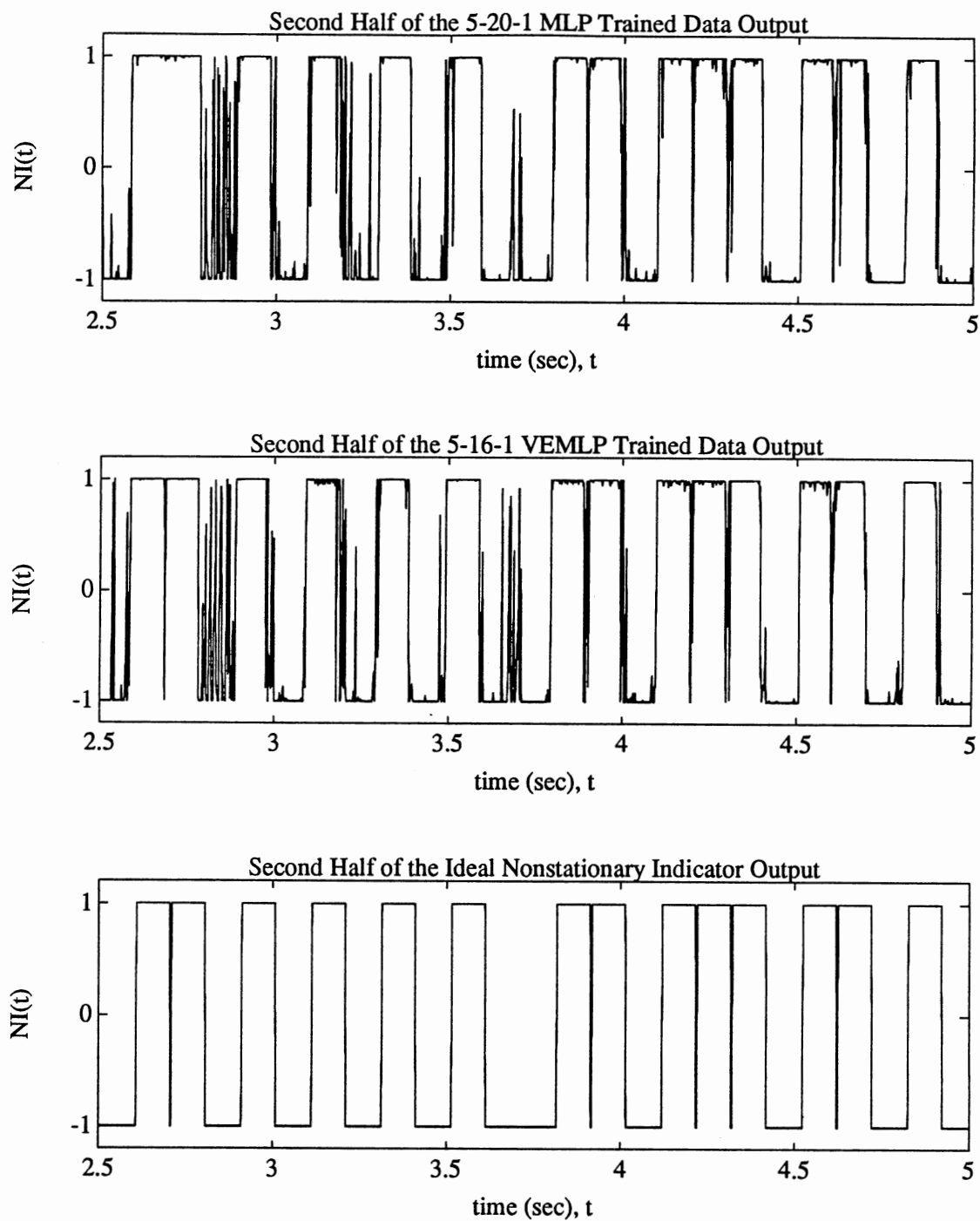


Figure V.21 The Nonstationary Indicator Output of a 5-20-1 Multilayer Perceptron and a 5-16-1 Volterra Expansion Multilayer Perceptron on the Trained Data.

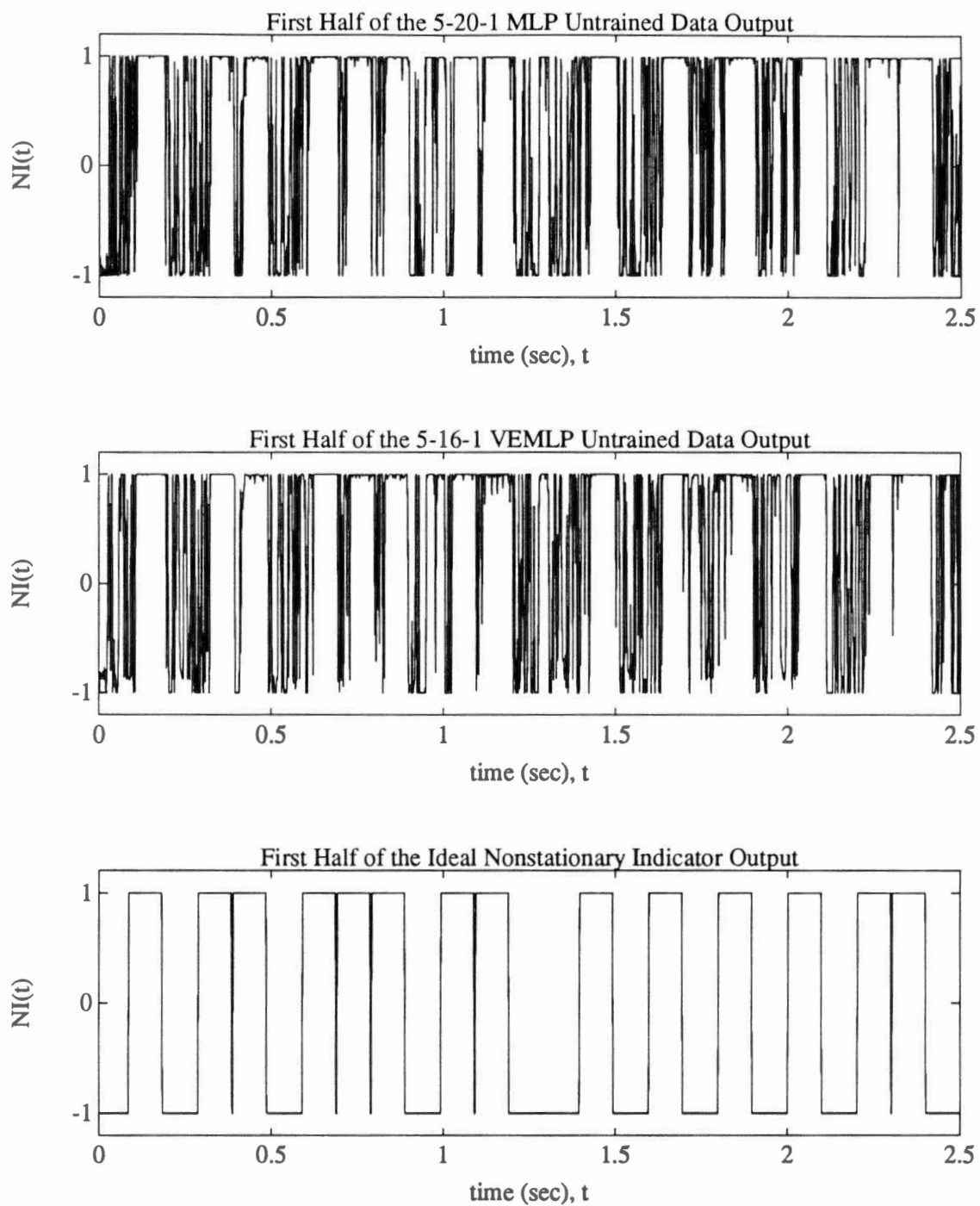


Figure V.22 The Nonstationary Indicator Output of a 5-20-1 Multilayer Perceptron and a 5-16-1 Volterra Expansion Multilayer Perceptron on Untrained Data.

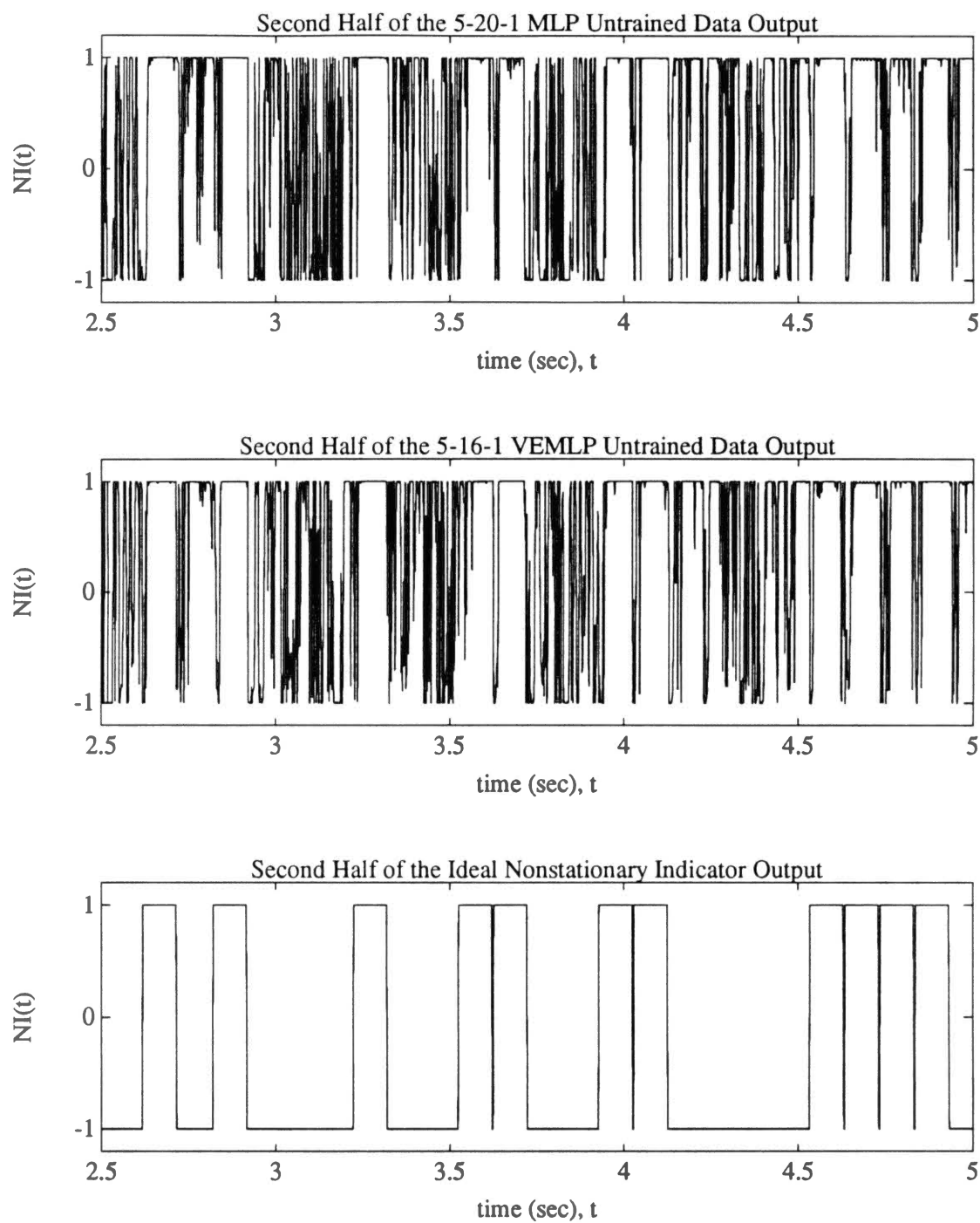


Figure V.23 The Nonstationary Indicator Output of a 5-20-1 Multilayer Perceptron and a 5-16-1 Volterra Expansion Multilayer Perceptron on Untrained Data.

characterize a general relationship between the Wigner distribution features and the nonstationary indicator output. The reader should note that the untrained data was just the trained data reversed in time. A time reversed copy of the trained data was used because the results for actual untrained data are significantly worse and the time reversed data allowed for easier comparisons.

Because the neural network implementation of the Wigner distribution feature based nonstationary indicator does not function well for the very simple case, the nonstationary indicator does not seem plausible. In fact, other training signals were simulated and trained with even worse results occurring. Thus, one of two things is wrong: either the multilayer perceptrons are not mapping the Wigner distribution features to the nonstationary indicator output, or the Wigner distribution features do not contain a general relationship between the Wigner distribution features and the nonstationary indicator. Because the multilayer perceptron and the Volterra expansion multilayer perceptron function properly for other nonlinear mappings, the conclusion is that Wigner distribution features do not allow a nonstationary indicator implementation.

Because a nonstationary indicator does not appear implementable using Wigner distribution features, what can be said about the time varying spectral energy information in the Wigner distribution features? The nonstationary indicator was proposed specifically to investigate whether the Wigner distribution features actually contain time varying spectral energy information. Because a function that indicates the amount a signal is stationary or nonstationary is much simpler than actual time varying spectral estimates, the nonstationary indicator should be quite simple for a time-frequency spectral energy to produce. However, the examples executed for this thesis indicate that time-frequency spectral energy moments do not provide enough information to compute a nonstationary indicator. Thus, other time-frequency spectral energy features may provide the information necessary for implementing a robust nonstationary indicator, although none could be found. Therefore, although the Wigner distribution features, defined in the thesis, are

well conceived (since the instantaneous power and instantaneous frequency are among them), the significant amount of time varying information indicated in the literature does not seem available, at least for the nonstationary indicator.

The lack of time-frequency spectral energy information in the moment time-frequency spectral energy features can probably be attributed to the fact that the features are very noisy. This noise can be seen in the examples provided in Chapter III. The noisy features occur since the time-frequency representations are themselves noisy. The reader should note that several different versions of filtered Wigner distribution features were tested on the nonstationary indicator implementation, but no positive results occurred. Therefore, the Wigner distribution features defined in this thesis are not very useful for directly implementing the nonstationary indicator. However, a usable nonstationary indicator may be possible if additional signal processing is performed on the nonstationary indicator output. The reason better performance is possible results from the observation that the nonstationary indicators appears more noisy for nonstationary signals. Performance may also be increased by modifying the neural network training algorithm to force lower errors on nonstationary signals.

Chapter Summary

This chapter has covered the proposed nonstationary indicator in context of time-frequency spectral energies. Specifically, the nonstationary indicator contained time-frequency based features and a method of mapping these features to the nonstationary indicator output. As for the time-frequency features, a thorough review of two-dimensional pdf moments was given so that a set of time-frequency features could be defined. These statistical based moments were defined and were applied to the Wigner distribution. Next, examples were provided to illustrate the performance of an ideal nonstationary indicator. An extensive review of the multilayer perceptron was also given because the multilayer perceptron implements the nonstationary indicator's mapping. In

addition to the basic multilayer perceptron review, a new neural network, the Volterra expansion multilayer perceptron, was defined and discussed. Although versions of the Volterra expansion multilayer perceptron were found in a couple of obscure papers, proofs for the update rules were never presented. Thus, the Volterra expansion multilayer perceptron discussion is considered original material. After the multilayer perceptron and the Volterra expansion multilayer perceptron software was validated, modifying the Wigner distribution features to contain more time-frequency spectral energy information was discussed. Finally, the entire chapter's work was applied to a very simple application of the nonstationary indicator. Unfortunately for the time-frequency spectral energy, the example indicates that the Wigner distribution features do not contain enough time varying spectral information to directly implement the nonstationary indicator.

Because Chapter IV indicated that time-frequency spectral energies cannot replace the spectrogram for estimating time varying spectral energy, time-frequency spectral energies' usefulness rested with the hope that features computed on time-frequency representations allow new or simpler signal processing algorithms. As such, a relatively simple signal processing application, the nonstationary indicator, was proposed to investigate whether time-frequency spectral energy features contained useful time varying spectral energy information. In order to give the time-frequency spectral energy nonstationary indicator the best chance for success, a multilayer perceptron mapped the time-frequency spectral energy features to the nonstationary indicator output. Because the multilayer perceptron can arbitrarily map any input to any output, success or failure of the nonstationary indicator resided with the amount of time varying spectral energy information contained in the time-frequency spectral energy features. Because the nonstationary indicator could not be made to function robustly on a simple example, a conclusion was made that the time-frequency spectral energy features do not contain enough useful time varying spectral information to implement the nonstationary indica-

tor. However, including additional signal processing or modifying the learning algorithm may allow implementing a nonstationary indicator. Because the nonstationary indicator does not encompass all signal processing problems, no general conclusion can be made concerning the amount of time varying spectral information contained in time-frequency spectral energies. Possibly other Wigner distribution features could have implemented the nonstationary indicator.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Summary

The main goal of this thesis was to investigate techniques of estimating spectral energies for time varying or nonstationary signals. The most popular time varying spectral energy method is the spectrogram which intuitively extends the Fourier transformation energy by windowing time varying signals. Although the spectrogram performs well on short-time time-invariant signals, the spectrogram produces poor spectral energy estimates for highly nonstationary signals or during signal transition regions. Thus, alternate methods for estimating time varying spectral energies have been and currently are heavily researched. The previous research divides into three distinct methods.

The first method attempts to extend the Fourier transformation to obtain spectral energies that depend on time. The spectrogram is the most popular method for extending the Fourier transformation. The second method describes spectral energies in terms of a time-frequency representation which corresponds to a two-dimensional pdf. A time-frequency representation indicates how a signal's energy distributes over time and frequency. One advantage of time frequency representations is that the marginal distributions typically equal useful signal processing quantities. Currently, the most popular time-frequency representation being investigated is the Wigner distribution. The

third method replaces the Fourier transformation's complex exponential basis function with functions specifically designed for time varying signals. The currently popular wavelet transformation comprises the bulk of this method's research.

Chapter II reviewed the historically significant and currently popular techniques generated by these three methods. For each reviewed time varying spectral energy technique, the rationale for its existence and major properties of the technique are discussed. In addition to specific time varying spectral energy techniques, Cohen's generalized time-frequency representation (which allows studying various time varying spectral energies) is also discussed. Besides allowing some limited comparisons between different time varying spectral energies, Cohen's generalized time-frequency representation indicates that the Wigner distribution generates many other time varying spectral energies. Thus, further understanding the Wigner distribution may provide valuable insights into other time varying spectral energies. Therefore, Chapter III thoroughly investigated the Wigner distribution.

Although Chapter III presented many Wigner distribution theoretical properties, the main goal of Chapter III was to compare the Wigner distribution to the spectrogram. The Wigner distribution was compared to the spectrogram because the spectrogram is the most popular method of estimating time varying spectral energies. The comparison between the Wigner distribution and the spectrogram was presented mainly with examples. These examples indicated that the Wigner distribution produces better spectral energies for chirp signals and similar spectral energies for signals containing only one sinusoid. However, the Wigner distribution for signals containing more than one sinusoid contain two major problems when compared to the spectrogram. First, large negative values exist which disallow interpreting the Wigner distribution as a two-dimensional pdf. Second, spurious spectral energies, called crossterms, occur for every pair of frequencies contained by the signal. Thus, the Wigner distribution cannot replace the spectrogram for estimating time varying spectral energies. Not all researchers

agree with this result because some researchers filter the Wigner distribution in an attempt to remove negative values and crossterms. However, Wigner distribution filtering typically nullifies many of the properties that make implementing the Wigner distribution desirable.

On the positive side, Chapter III showed that the Wigner distribution may be useful in signal processing. Several examples compared features generated by the Wigner distribution and the spectrogram. These comparisons indicated that the Wigner distribution may provide more time varying signal energy information than the spectrogram. Because time varying spectral information is not the same as the time varying spectral energy itself, the main thesis goal (analyzing time varying spectral energies) divides into two sub-goals. The first sub-goal investigated whether any techniques exist that are capable of replacing the spectrogram for estimating robust time varying spectral energies. The second sub-goal investigated whether features computed on time varying spectral energies produce better information than spectrogram features. Chapters IV and V investigate these two sub-goals separately.

Chapter IV addressed the first thesis sub-goal which asks whether any alternate time varying spectral energy can replace the spectrogram. In order to answer this question, a new technique, called the correlation approach to time-frequency representations, was developed to allow comparing potential time varying spectral energies more easily than Cohen's generalized time-frequency representation. This is one of the main contributions of this research. The correlation approach extended the idea that spectral energy can be computed as the Fourier transformation of an autocorrelation function when the underlying random process is stationary. The extension was performed by defining three new quantities: the time-frequency autocorrelation function, the time-frequency expected value, and the time-frequency spectral energy.

The time-frequency autocorrelation function,

$$A_x(t_1, t_2) \stackrel{\Delta}{=} E_A[x(t_1)x^*(t_2)], \quad (\text{VI.1})$$

is identical to a nonstationary autocorrelation function that depends on two time variables except the time-frequency expected value, $E_A[x(t_1)x^*(t_2)]$, replaces the expected value.

The time-frequency expected value is similar to an ordinary expected value except that no assumption is made concerning the underlying random process's classification.

Because defining the time-frequency expected value is arbitrary, the time-frequency expected value can be estimated by time averages, ensemble averages, or any another conceivable average. Arbitrarily specifying the time-frequency expected value differs significantly from the normal expected value which is properly estimated using time averages when the random process is ergodic and properly estimated using ensemble averages when the random process is stationary. This latitude allowed describing many different time-frequency autocorrelation functions without concern of whether the random process is stationary or nonstationary.

The correlation approach was completed by defining the time-frequency spectral energy as

$$TF_x(t, f) \stackrel{\Delta}{=} \int_{-\infty}^{\infty} A_x(t, \tau) e^{-i2\pi f\tau} d\tau, \quad (\text{VI.2})$$

which is the Fourier transformation of the time-frequency autocorrelation function with respect to a time difference variable. Because the Fourier transformation is with respect to a time difference variable, the two time variable time-frequency autocorrelation function required rewriting. The time-frequency autocorrelation function was rewritten by transforming the two time variables to a time variable and a time difference variable. Thus, the time-frequency spectral energy converts the time-frequency autocorrelation function into a function of time and frequency. Because the time-frequency spectral

energy requires transforming the time-frequency autocorrelation function variables, the time-frequency spectral energy depends on two things: the method chosen for computing the time-frequency expected value and the two time variable transformations.

The first step in applying the correlation approach to analyzing time varying spectral energies was to write the common time-frequency representations in terms of the time-frequency spectral energy. This first step produced three basic ways to compute the time-frequency expected value: the time average, the ensemble average, and the one sample function ensemble average. The reader should note that because ensemble averages require averaging over an infinite number of sample functions, the ensemble average and one sample function ensemble average differ significantly. Using these three time-frequency expected values and the most obvious transformation

$$\begin{array}{lcl} t = t_1 & & t_1 = t \\ \tau = t_1 - t_2 & \Leftrightarrow & t_2 = t - \tau \end{array} \quad (VI.3)$$

the time-frequency spectral energy produces the periodogram, the power spectral density, and the Rihaczek distribution. By changing the time variable transformation (VI.3) to

$$\begin{array}{lcl} t = \frac{t_1 + t_2}{2} & & t_1 = t + \frac{\tau}{2} \\ \tau = t_1 - t_2 & \Leftrightarrow & t_2 = t - \frac{\tau}{2} \end{array} \quad (VI.4)$$

the time average and the ensemble average time-frequency spectral energies remain the same, but the one sample function ensemble average time-frequency spectral energy becomes the Wigner distribution. Thus, the only difference between the Rihaczek distribution and the Wigner distribution is a slight change in the time variable transformation. However, the Rihaczek distribution and the Wigner distribution differ significantly because the Rihaczek distribution is complex valued while the Wigner distribution is real valued.

After showing how the time-frequency spectral energy produces the periodogram, the power spectral density, and the Wigner distribution, several properties were derived for the correlation approach. First, the time-frequency spectral energy is real valued if the time-frequency autocorrelation function is an even function with respect to the time difference variable. This fact implies that the Wigner distribution is real valued and the Rihaczek distribution is complex valued. Additionally, all time average time-frequency autocorrelation functions produce real valued time-frequency spectral energies. However, when using ensemble average time-frequency expected values, transformation (VI.3) based time-frequency autocorrelation functions may not produce real valued time-frequency spectral energies, even though the time-frequency autocorrelation function based upon transformation (VI.4) produces real time-frequency spectral energies. This result is interesting because current autocorrelation function spectral energy methods implicitly depend on transformation (VI.3). Since time-frequency autocorrelation functions based upon transformation (VI.3) may not produce real valued spectral energies, current autocorrelation function based spectral energy methods should be modified to use transformation (VI.4).

The next set of properties from Chapter IV determined when a time-frequency spectral energy contains the instantaneous power and the power spectral density as marginal distributions. These properties indicate when time-frequency spectral energies potentially produce useful features. For a time-frequency spectral energy to contain the instantaneous power, the time-frequency autocorrelation function (with the time difference variable equalling zero) must equal the instantaneous power. A time-frequency spectral energy produces the power spectral density if the time-frequency autocorrelation function uses ensemble averages and linear time variable transformations. Using these properties and additional properties, Chapter IV showed that time average time-frequency expected value based time-frequency spectral energies cannot produce the instantaneous

power. However, ensemble average and one sample function ensemble average time-frequency expected value based time-frequency spectral energies can produce the instantaneous power.

The question still remained whether any new time varying spectral energy technique could replace the spectrogram. By defining a general transformation (IV.74), it was shown that real valued time-frequency spectral energies (capable of producing the instantaneous power) result only from time-frequency autocorrelation functions defined with transformation (VI.4) and ensemble averages when considering only linear transformations. Thus, the Wigner distribution is practically the only time-frequency spectral energy which is computable on nonstationary signals that contains desirable properties and is based upon linear transformations.

The correlation approach analysis continues by considering nonlinear time variable transformations. Even though nonlinear transformations are more difficult to work with than linear transformations, nonlinear transformations were considered due to the desire to replace the spectrogram. The one nonlinear transformation that seemed reasonable modified transformation (VI.3) so that the time-frequency autocorrelation function is always an even function with respect to the time difference variable. Several properties indicate that transformation

$$\begin{array}{ccc} t = t_1 & & t_1 = t \\ |\tau| = t_1 - t_2 & \Leftrightarrow & t_2 = t - |\tau| \end{array} \quad (\text{VI.5})$$

defines such a transformation. As with the other two transformations, this transformation produces the periodogram and the power spectral density for time average and ensemble average time-frequency expected value based time-frequency autocorrelation functions. However, the one sample function ensemble average time-frequency expected value based time-frequency autocorrelation function produces Turner's instantaneous power spectrum. In addition, properties derived for time-frequency spectral energies based upon (VI.5) closely resemble properties established for time-frequency spectral energies

that use transformation (VI.4). Therefore, Turner's instantaneous power spectrum acts like the Wigner distribution. The question that remains is whether Turner's instantaneous power spectrum estimates time varying spectral energies more robustly than the Wigner distribution. Chapter IV showed that, like the Wigner distribution, Turner's instantaneous power spectrum produces crossterms for signals that contain more than one sinusoid. Thus, Turner's instantaneous power spectrum performs no better than the Wigner distribution. The correlation approach analysis of time varying spectral energy techniques stopped here because no other useful transformation or time-frequency expected value computing method could be found.

Thus, Chapter IV answered the question posed by this thesis' first sub-goal. The question was whether any alternate time-frequency spectral energy can replace the spectrogram. The correlation approach answered this question by concluding that no time-frequency spectral energy seems capable of replacing the spectrogram. Although many more time-frequency spectral energies exist using different transformations and different time-frequency expected values, all the usable time-frequency spectral energies that one could come up with were presented in Chapter IV. Because no time-frequency spectral energy can replace the spectrogram, time-frequency spectral energies may be useful in obtaining time varying spectral energy information. Chapter V investigated this question.

Chapter V addressed the second thesis sub-goal that asks whether features computed on time-frequency spectral energies produce better time varying spectral energy information than spectrogram features. To answer this question, two separate issues were investigated. First, what time-frequency spectral energy features should be used, and second, how should the features be tested?

Because many researchers have claimed that the first two moments of the Wigner distribution contain time varying spectral information, the feature selection began with these two moments. However, to understand what information these two moments con-

tained, Chapter V reviewed moment based features for generic two-dimensional pdfs. This general moment discussion produced five Wigner distribution features: the instantaneous power, the instantaneous frequency, the variance, the skewness, and the kurtosis. The instantaneous power corresponds to the Wigner distribution's marginal distribution, and the instantaneous frequency represents the Wigner distribution's conditional mean. The last three Wigner distribution features correspond to the Wigner distribution's second, third, and fourth order conditional central order moments.

Of the five proposed Wigner distribution moments, only the instantaneous power and the instantaneous frequency have been widely acclaimed to contain time varying spectral energy information. In addition, the Wigner distribution variance has also been previously defined [Cla83] but not thoroughly tested. The first three Wigner distribution features correspond to statistical quantities that are usually computed to obtain distribution information. However, the marginal distribution, the mean, and the variance contain complete statistical information only if a two-dimensional pdf is Gaussian. Chapter V illustrated that the Wigner distribution is not Gaussian distributed. As such, additional information is available with higher order moments. Thus, the skewness and kurtosis were introduced as Wigner distribution features.

Unfortunately, the Wigner distribution is not an actual two-dimensional pdf, so the features cannot be directly implemented using the moment formulae. Instead, two potential numerical problems must be addressed. First, all five Wigner distribution features, excluding the instantaneous power, contain divisions that may not be defined. Although the divisors rarely equal zero, values near zero occur often enough to cause erratic behavior in these features. To counteract the near zero divisions, a median filter is used to filter the features. Several experiments indicated that the median filter reduced the problem while not significantly effecting the actual signal. The second potential numerical problem occurs in the Wigner distribution skewness and kurtosis where the square root of the Wigner distribution variance is computed. Because the Wigner distribution variance may

be negative, the absolute value of the Wigner distribution variance is used instead. Thus, these two simple modifications of the moment formulae allow applying the moment formulae to the Wigner distribution.

Chapter V then attempted to investigate whether the Wigner distribution features actually contain time varying spectral information. The best method for evaluating whether the Wigner distribution features contained time varying spectral information was to apply the features to some signal processing application. For the signal processing application, a new function called the nonstationary indicator was introduced. Using the Wigner distribution features as input, the nonstationary indicator's output states whether a signal is stationary or nonstationary. Because such an application is not easily implemented using standard signal processing techniques and because the Wigner distribution features supposedly contains time varying spectral energy information, the nonstationary indicator made a good evaluation platform. A problem remains as to how the nonstationary indicator takes the Wigner distribution feature input and produces the signal classification (defined as a 1.0 for stationary signals and a -1.0 for nonstationary signals).

Because the nonstationary indicator can be described as a mapping which takes a set of numbers (the Wigner distribution features) to a real number (1 or -1), a technique was needed that allowed the Wigner distribution features a good chance for success. Thus, the nonstationary indicator was implemented using a multilayer perceptron neural network which is capable of learning arbitrary mappings of real numbers to real numbers. Another advantage of using a multilayer perceptron is that by not allowing the number of weights to exceed $1/5$ of the number of training samples, the multilayer perceptron generalizes the relationship between the input and the output. Thus, if the multilayer perceptron fails to converge or the final weights perform poorly on untrained data, then no robust relationship exists between the input and the output. In addition, if any of the Wigner distribution features fail to produce information necessary for robust

nonstationary indicator operation, then the multilayer perceptron will learn to ignore the poor features. Therefore, the multilayer perceptron was a reasonable method for implementing the nonstationary indicator.

During the course of implementing the nonstationary indicator, a new neural network was developed that may allow easier generalizations. This new neural network, the Volterra expansion multilayer perceptron, was loosely based upon the Volterra expansion used in nonlinear system theory. Although some research papers have mentioned this type of network before and even provided training formulae, no proofs for the training rules have been provided. In addition, few justifying examples have been published. Chapter V not only presented training formulae proofs but also presented an example where the Volterra expansion multilayer perceptron trained faster than a comparable multilayer perceptron. Thus, the Volterra expansion multilayer perceptron also implements the nonstationary indicator.

Recall that the purpose of the nonstationary indicator was to evaluate whether time varying spectral energy information exists in the Wigner distribution features. Thus, a final production version of the nonstationary indicator is not desirable but only a decision as to whether the Wigner distribution features can implement a nonstationary indicator. As such, the experiments began on a simple signal that consisted of 25 different sinusoids and 25 different chirp signals that were randomly selected. Several varieties of Volterra expansion multilayer perceptron and multilayer perceptron configurations were tested along with additional processing of the Wigner distribution features. However, nothing could force a robust nonstationary implementation. Because the Volterra expansion multilayer perceptron and the multilayer perceptron are capable of arbitrary mappings, the fault was the input to the networks' input. Thus, the Wigner distribution features do not provide enough information to implement the nonstationary indicator. Therefore, the Wigner distribution features do not contain useful time varying spectral energy information in the context of this example. Because the other time-frequency spectral energies

produced similar features (in fact they all produce the instantaneous power), the time-frequency spectral energies discussed in this thesis do not contain enough time varying spectral energy information to implement the nonstationary indicator. Thus, it seems that the time varying spectral energy information contained in time-frequency spectral energies is either not in the proper form to implement the nonstationary indicator or other features capture the nonstationary indicator more effectively.

The thesis results can be summarized by stating first that time-frequency spectral energies do not seem capable of replacing the spectrogram. Next, this thesis showed that, in general, features computed on time-frequency spectral energies do not contain enough time varying spectral information to robustly implement the nonstationary indicator. Combining these two conclusions forms the overall thesis conclusion that time-frequency spectral energies do not seem to produce any more time varying spectral energy information than produced by the spectrogram.

Not only did this thesis produce conclusions concerning the usefulness of time varying spectral energies, but several interesting things resulted from the main study. First, the correlation approach to time-frequency representations was developed and was shown useful in analyzing the time-frequency spectral energies. Additionally, the correlation approach can be used to analyze future methods of time-frequency spectral energies. Second, the fact that higher order moments provide additional information than what is produced by just the marginal distribution, the mean value, and the variance for non-Gaussian signals was introduced. Although these higher order moments did not help the Wigner distribution or the other time-frequency spectral energies discussed in this thesis, future time varying spectral energy techniques may be able to use this information. Third, the Volterra expansion multilayer perceptron was introduced as an alternative to the regular multilayer perceptron. Not only were the Volterra expansion multilayer per-

ceptron's training rules proved, but the Volterra expansion multilayer perceptron was shown capable of speeding up training time compared to a comparable multilayer perceptron for an example problem.

Future Work

This thesis' main goal was to analyze alternate techniques for estimating time varying spectral energies and for obtaining information from time varying spectral energies. To a certain degree, this thesis has met this goal by concluding that the studied time-frequency spectral energies cannot replace the spectrogram and do not produce features that contain enough time varying spectral energy information to implement a simple problem. However, notice that not every possible time-frequency spectral energy was tested. In fact, only nine main time-frequency spectral energies were heavily investigated. The reason for studying only nine time-frequency spectral energies was that the three main time-frequency expected values and three main time variable transformations used were the only obvious choices. Other time-frequency expected values and transformations may exist, so the respective time-frequency spectral energies may contain useful time varying spectral energy information. Investigating alternate transformations probably would be the best approach for developing new time-frequency spectral energies, especially looking at nonlinear transformations. As for new time-frequency expected values, different definitions exist, but new time-frequency expected values would most likely be a radical departure from what is generally accepted as an expected value. Nonetheless, investigating new transformations and new time-frequency expected values remains a wide open research area that might produce interesting results.

In addition to new transformations and time-frequency expected values, the correlation approach also has future research potential in investigating new time varying spectral energy techniques. No reason exists to expect that new methods cannot be investigated

in terms of the time-frequency spectral energy because seemingly unrelated techniques (the Wigner distribution, the Rihaczek distribution, and Turner's instantaneous power spectrum) are easily related to the time-frequency spectral energy.

Other future works could include alternate implementations of the nonstationary indicator. One particular alternate implementation would modify the neural network training algorithm to reduce the extra nonstationary indicator noise found for nonstationary signals. In addition, using time-frequency spectral energy features for time varying spectral energy information is the beginning of new research that will develop radical new techniques designed specifically for this purpose. A radical new approach would dispense of the complex exponential basis function found in the time-frequency spectral energy and replace it with some function specifically designed for feature extraction. This approach could produce very useful results because the old method, which used features computed on time-frequency spectral energies designed specifically to estimate time varying spectral energies, was not designed to produce features. An example of a radical departure is the wavelet transformation which is currently under heavy investigation. Because the wavelet transformation is general, wavelet functions may be designed specifically to compute features.

The final new research method that this thesis has produced is the Volterra expansion multilayer perceptron. Although including the Volterra expansion into the multilayer perceptron did not significantly increase system complexity, the resulting decrease in training time for the nonlinear function simulated was quite favorable. A significant amount of research remains regarding exactly how the Volterra expansion multilayer perceptron operates compared to the normal multilayer perceptron. In addition, no mention was made in this thesis concerning higher order Volterra expansions. The possibility exists to build even more powerful networks if third or fourth order Volterra expansions are included. Thus, applying the Volterra expansion to the multilayer perceptron is an open research area.

Overall, the most promising research direction involves investigating new time-frequency spectral energy basis functions, specifically designed for extracting time varying spectral energy information. The wavelet transformation is a good start, but other non-orthogonal basis functions could be used. Once new features are computed, the new Volterra expansion multilayer perceptrons may allow for better understanding of how these new features apply to signal processing.

REFERENCES

- [Abe89] R. Abeysekera and B. Boashash, "Time-Frequency Domain Features of ECG Signals: Their Application in P-Wave Detection Using the Cross Wigner-Ville Distribution," in *Proc. IEEE ICASSP-89*, vol. 12, pp. 1524-1527, 1989.
- [Ben86] J. Bendat and A. Piersol, *Random Data: Analysis and Measurement Procedures*, John Wiley and Sons, New York, 1986.
- [Boa87] B. Boashash and P. Black, "An Efficient Real-Time Implementation of the Wigner-Ville Distribution," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 35, pp. 1611-1618, 1987.
- [Boa88] B. Boashash, "Note on the Use of the Wigner Distribution for Time-Frequency Signal Analysis," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 36, pp. 1518-1521, 1988.
- [Bou83] G. Bourdueux-Bartels, *Time-Frequency Signal Processing Algorithms: Analysis and Synthesis using Wigner Distributions*, Ph.d. dissertation, Rice University, Houston, Texas, 1983.
- [Buc78] R. Buck, *Advanced Calculus*, McGraw-Hill Book Company, New York, New York, 1978.
- [Che83] D. Chester, F. Taylor, and M. Doyle, "Applications of the Wigner Distribution to Speech Processing," in *Proc. IEEE ICASSP-83*, pp. 98-102, 1983.
- [Che84] C. Chen, *Linear System Theory and Design*, Holt, Rinehart and Winston, New York, 1984.
- [Cho89] H. Choi and W. Williams, "Improved Time-Frequency Representations of Multicomponent Signals Using Exponential Kernalns," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 37, pp. 862-871, 1989.
- [Cla80a] T. Claasen and W. Mecklenbrauker, "The Wigner Distribution - A Tool for Time-Frequency Signal Analysis - Part I Continuous - Time Signals," *Philips Journal of Research*, vol. 35, pp. 212-250, 1980.

- [Cla80b] T. Claasen and W. Mecklenbrauker, "The Wigner Distribution - A Tool for Time-Frequency Signal Analysis - Part II: Discrete - Time Signals," *Phillips Journal of Research*, vol. 35, pp. 276-300, 1980.
- [Cla83] T. Claasen and W. Mecklenbrauker, "The Aliasing Problem in Discrete-Time Wigner Distributions," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 31, pp. 1067-1072, 1983.
- [Coh66] L. Cohen, "Generalized phase-space distribution functions," *Journal of Mathematical Physics*, vol. 7, pp. 781-786, 1966.
- [Coh89] L. Cohen, "Time-Frequency Distributions - A Review," *Proceedings of the IEEE*, vol. 77, pp. 941-981, 1989.
- [Cot90] N. Cotter, "The Stone-Weierstrass Theorem and Its Application to Neural Networks," *IEEE Transactions on Neural Networks*, vol. 1, pp. 290-295, 1990.
- [Day84] D. Day and R. Yarlagadda, "The Modified Discrete Wigner Distribution and Its Application to Acoustic Well Logging," *Acoustical Imaging*, vol. 13, pp. 293-313, 1984.
- [Day87] D. Day, *The Modified Wigner Distribution and Two Dimensional Allan Variance with Applications to Acoustic Well Logs and Speech*, Ph.d. dissertation, Oklahoma State University, Stillwater, Oklahoma, 1987.
- [Egg90] J. Eggermont and G. Smith, "Characterizing auditory neurons using the Wigner and Rihaczek distributions: A comparison," *The Journal of the Acoustical Society of America*, vol. 87, pp. 246-259, 1990.
- [Fan50] R. Fano, "Short-Time Autocorrelation Functions and Power Spectra," *Journal of the Acoustical Society of America*, vol. 22, pp. 546-550, 1950.
- [Fla84] P. Flandrin and B. Escudie, "An Interpretation of the Pseudo-Wigner-Ville Distribution," *Signal Processing*, vol. 16, pp. 27-36, 1984.
- [Gab46] D. Gabor, "Theory of Communication," *Institute of Electrical Engineering Part III*, vol. 93, pp. 429-457, 1946.
- [Gar86] W. Gardner, *Introduction to Random Processes With Applications to Signals and Systems*, Macmillan, New York, 1986.

- [Gei88] C. Geisler, "Representation of speech sounds in the auditory nerve," *Journal of Phonetics*, vol. 16, pp. 19-35, 1988.
- [Gul89] W. Gulick, G. Gescheider, and R. Frisna, *Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics*, Oxford University Press, New York, 1989.
- [Hec87] R. Hecht-Nielsen, "Kolmogorov's Mapping Neural Network Existence Theorem," in *Proc. IEEE ICNN - 87*, pp. III-11-13, 1987.
- [Hec89] R. Hecht-Nielsen, "Theory of the Backpropagation Neural Network," in *Proc. IEEE ICNN-89*, pp. I-593-605, 1989.
- [Kli91] C. Klimasauskas, "Applying Neural Networks: Part III," *PC AI*, pp. 20-24, 1991.
- [Koe46] W. Koenig, H. Dunn, and L. Lacy, "The Sound Spectrograph," *Journal of the Acoustical Society of America*, vol. 18, pp. 19-54, 1946.
- [Kol57] A. Kolmogorov, "On the representation of continuous functions of several variables by superposition of continuous functions of one variable and addition," *Dokl. Akad. Nauk SSSR*, vol. 114, pp. 369-373, 1957.
- [Kum84] B. Kumar and C. Carroll, "Performance of Wigner distribution function based detection methods," *Optical Engineering*, vol. 23, pp. 732-737, 1984.
- [Lev64] M. Levin, "Instantaneous Spectra and Ambiguity Functions," *IEEE Transactions on Information Theory*, vol. IT-10, pp. 95-97, 1964.
- [Lip87] R. Lippman, "An Introduction to Computing with Neural Networks," *IEEE AASP Magazine*, pp. 4.22, 1987.
- [Lor76] G. Lorentz, "The 13th Problem of Hilbert," in F.E. Browder (Ed.), *Mathematical Developments Arising from Hilbert Problems*, American Mathematical Society, Providence, R.I., 1976.
- [Mal90] D. Malkoff and L. Cohen, "A Neural Network Approach to the Detection Problem Using Joint Time-Frequency Distributions," in *Proc. IEEE ICASSP-90*, vol. 1, pp. 2739-2742, 1990.
- [Mil81] G. Miller, *Language and Speech*, W.H. Freeman and Company, San Francisco, 1981.

- [Osh87] D. O'Shaughnessy, *Speech Communication, Human and Machine*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1987.
- [Osh90] P. O'Shea and B. Boashash, "Instantaneous Frequency Estimation Using the Cross Wigner-Ville Distribution with Applications to Nonstationary Transient Detection," in *Proc. ICASSP-90*, vol. 4, pp. 2887-2890, 1990.
- [Pag52] Page, "Instantaneous Power Spectra," *Journal of Applied Physics*, vol. 23, pp. 103-106, 1952.
- [Pao89] Y. Pao, *Adaptive Pattern Recognition & Neural Networks*, Addison-Wesley Publishing Company, Inc., Reading Massachusetts, 1989.
- [Pap62] A. Papoulis, *The Fourier Integral and its Applications*, McGraw-Hill Book Company, New York, 1962.
- [Pap84] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Book Company, New York, 1984.
- [Pis85] D. Pisoni, "Speech perception: some new directions in research and theory," *Journal of the Acoustical Society of America*, vol. 78(1), pp. 381-388, 1985.
- [Pol88] H. Pollard, "Feature Analysis of Musical Sounds," *Acustica*, vol. 65, pp. 232-244, 1988.
- [Pre86] W. Press, B. Flannery, S. Tevkolsky, and W. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1986.
- [Pre87] D. Preis, F. Halwatsch, P. Bloom, and J. Deer, "Wigner Distribution Analysis of Filters with Perceptible Phase Distortion," *Journal of the Audio Engineering Society*, vol. 35, pp. 1004-1012, 1987.
- [Rab78] L. Rabiner, and R. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1978.
- [Rih68] A. Rihaczek, "Signal Energy Distribution in Time and Frequency," *IEEE Transactions on Information Theory*, vol. IT-14, pp. 369-374, 1968.
- [Ril89] M. Riley, *Speech Time-Frequency Representations*, Kluwer Academic Publishers, Norwell, Massachusetts, 1989.
- [Ros62] F. Rosenblatt, *Principles of Neurodynamics*. Spartan Books, Washington, DC, 1962.

- [Rum81] D. Rumelhart, G. Hilton, and R. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1. Foundations*, edited by D.E. Rumelhart and J.L. McClelland (MIT, Cambridge, MA), pp. 318-362, 1981.
- [Sch62] M. Schroeder and B. Atal, "Generalized Short-Time Power Spectra and Autocorrelation Functions," *Journal of the Acoustical Society of America*, vol. 34, pp. 1679-1683, 1962.
- [Sch80] M. Schwartz, *Information Transmission, Modulation, and Noise*, McGraw-Hill, New York, 1980.
- [Sch86] E. Schwab and H. Nusbaum, *Pattern Recognition by Humans and Machines, Volume I: Speech Perception*, Academic Press, San Diego, California, 1986.
- [Scz80] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, John Wiley and Sons, New York, New York, 1980.
- [Spr65] D. Sprecher, "On the structure of continuous functions of several variables," *Transactions of the American Mathematical Society*, vol. 115, pp. 340-355, 1965.
- [Ste46] J. Steinberg and N. French, "The Portrayal of Visible Speech," *Journal of the Acoustical Society of America*, vol. 18, pp. 4-18, 1946.
- [Sun89] M. Sun, C. li, L. Sekhar, and R. Sclabassi, "Elimination of Cross-Components of the Discrete Pseudo-Wigner Distribution via Image Processing," in *Proc. IEEE ICASSP-89*, vol. 4, pp. 2230-2233, 1989.
- [Szu84] H. Szu and H. Caulfield, "The Mutual Time-Frequency Content of Two Signals," *Proceedings of the IEEE*, vol. 72, pp. 902-908, 1984.
- [Szu86] H. Szu, "Three Layers of vector outer product neural networks for optical pattern recognition," in *Proc. SPIE: Optical and Hybrid Computing*, vol. 634, pp. 312-330, 1986.
- [Tau74] T. Tau and R. Gonzalez, *Pattern Recognition Principles*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1974.
- [Tur54] C. Turner, "On the Concept of an Instantaneous Power Spectrum, and its Relationship to the Autocorrelation Function," *Journal of Applied Physics*, vol. 25, pp. 1347-1351, 1954.

- [Usu90] S. Usui and H. Araki, "Wigner Distribution Analysis of BSPM for Optimal Sampling," *IEEE Engineering in Medicine and Biology Magazine*, vol. 37, pp. 29-32, 1990.
- [Vel89] E. Velez and R. Absher, "Transient Analysis of Speech Signals Using the Wigner Time-Frequency Representation," in *Proc. IEEE ICASSP-89*, 1989.
- [Vel90] E. Velez and R. Absher, "Spectral Estimation based on the Wigner-Ville Representation," *Signal Processing*, vol. 20, pp. 325-346, 1990.
- [Vil48] J. Ville, "Theorie et Applications de la Notion de Signal Analitique," *Cables et Transmission*, vol. 2a, pp. 61-74, 1948.
- [Wid60] B. Widrow and M. Hoff, "Adaptive Switching Circuits," *1960 IRE WESCON Conv. Record, Part 4*, pp. 96-1-4, 1960.
- [Wid85] B. Widrow and S. Stearns, *Adaptive Signal Processing*, Prentice-Hall, New Jersey, 1985.
- [Wig32] E. Wigner, "On the Quantum Correction for Thermodynamic Equilibrium," *Physics Review*, vol. 40, pp. 749-759, 1932.
- [Woo53] P. Woodward, *Probability and Information Theory, with Applications to Radar*, McGraw-Hill Book Co. Inc., New York, 1953.
- [Yen87] N. Yen, "Time and frequency representation of acoustic signals by means of the Wigner distribution function: Implementation and interpretation," *Journal of the Acoustical Society of America*, vol. 81(b), pp. 1841-1850, 1987.
- [Yen90] N. Yen, L. Dragonette, and S. Numrich, "Time-frequency analysis of acoustic scattering from elastic objects," *Journal of the Acoustical Society of America*, vol. 87, pp. 2359-2370, 1990.
- [Yos85] W. Yost and D. Nielsen, *Fundamentals of Hearing, An Introduction*, Holt, Rinehart and Winston, New York, 1985.

APPENDIX

APPENDIX

A REVIEW OF THE HUMAN SPEECH AND HEARING SYSTEMS

Without thinking, most of us are able to speak what we want to communicate and easily understand other people's speech. Therefore, nature has provided us with a very robust system which easily accomplishes communication and reception tasks. Thus, incorporating as much of nature's system as possible into a signal processing system makes sense. Hence, when applying signal processing to the area of speech processing, the speech production system and the hearing system should be reviewed. As such, a short review of speech production modelling ideas is presented. Because the physical models describing hearing and speech understanding are not understood as well as the speech production system, a more physiological overview of the hearing system is presented to see if signal processing techniques can incorporate this knowledge.

A speech signal is transmitted through a medium (such as air) in the form of pressure waves that travel from a speaker's mouth to a listener's ear. This signal consists of pressure variations fluctuating with respect to time relative to a local atmospheric pressure. The fluctuating pressure is initiated by changes in air flow produced by the lungs, throat, and mouth [Osh87]. The speech production system can be decomposed into a source and a filter [Mil81]. The source corresponds to the source of changing air flow and is produced by the lungs and vocal cords. The lungs and vocal cords combine efforts to produce an airflow which fluctuates randomly (white noise) or to produce periodic puffs of air (an impulse train of air flow). The sounds produced by periodic

fluctuations are called voiced sounds. The frequency at which voiced fluctuations occur is termed the pitch period. An example of a voiced sound is the vowel /o/ found in the word boy. Notice when speaking the vowel /o/, how it sounds very periodic and has a very distinct tonality. Sounds produced by the randomly fluctuating inputs are called unvoiced sounds since the pitch period is not present. A typical unvoiced sound is the /s/ found in the word sound. Notice how the consonant /s/ sounds noisy compared the periodic sounding /o/.

The two input source airflows, produced by the lungs and vocal cords, are modified by the filter portion of the speech production model. The filter section includes the throat and mouth. The throat and mouth are considered filters because changes in their physical shape and length modify the spectral characteristics of the input source. A simple model for the throat is an open ended acoustical tube [Rab78]. The acoustical tube has the characteristic of resonating at frequencies associated with its quarter wavelength, three quarter wavelength, five quarter wavelength, etc. The acoustical tube contains anti-resonators at frequencies associated with its half wavelength, full wavelength, one and a half wavelength, etc. These resonances and anti-resonances indicate that depending on the length of the throat and mouth (the acoustical tube) a filter is defined. For a typical vowel, a resonance will occur approximately every 1000 Hz and likewise for the anti-resonances. Due to the limited amount of energy available by the input source, a speech signal's energy is contained in the frequency range of 0 Hz to 4000 Hz. Therefore, four resonances typically provide a majority of the filtering information of the throat. These resonant frequencies or formant frequencies are very important in modelling the speech production system. Adding slightly to the filter is the mouth which produces other filtering effects by changing the placement of the teeth and by providing different movements of the lips.

The physical speech production model description reviewed ideas used in speech coding and speech recognition algorithms. Since only four formant frequencies are typically prominent, a modelling scheme need only to incorporate the four formant frequencies into the model instead of attempting to describe the entire frequency spectrum. In fact, LPC algorithms define the speech production model as a system with 12 poles (eight poles for the four formants and four poles for the mouth) [Osh87]. Therefore, the entire spectral content of a stationary speech signal is reduced to 12 filter related parameters. As remarkable as it seems, this model contains enough of the original speech information to allow intelligible reconstruction if source type is correctly predicted. Because LPC analysis condenses so much information about a speech signal in so little space, the coefficients make very good features for a speech recognition system. However, in order to compute the filter parameters from the speech signal, the speech signal is assumed short-time time-invariant. Thus, the modelling techniques used in LPC analysis are not robust during periods of rapidly changing speech such as in the transition region between phonemes.

The major point in the speech production overview is that the spectral content of the speech signal is very important in speech processing algorithms. The fact that a few formant frequencies carry so much information can be deceiving, however, in the context of speech recognition. Although the formant information includes enough speech information for coding applications, the same may not necessarily be true for the recognition task. Other aspects of the speech signal may relay just as much information as the spectral content of the signal. The only way to find out is to investigate what operations are performed on a speech signal in the hearing system.

The hearing system is much more difficult to analyze than the speech production system. In the speech production system, an x-ray may be made of the throat and mouth to describe the three dimensional shape of the acoustical tube which describes the filter section of the system. It has been shown that acoustical analysis accurately describes

how different sounds are filtered [Osh87]. However, only a small portion of the hearing system is describable with acoustical science. The remaining parts of the hearing system are described in terms of physiological and neurological science. Adding more difficulty is the fact that there seem to be two different types of research performed in speech hearing. The first type of research investigates how the ear converts the pressure fluctuations into neurological signals. This research searches for speech signal features extracted by the hearing system. The other hearing research attempts to describe the information the brain uses for recognizing spoken speech. The most useful information for a speech recognition system would be knowledge of what speech characteristics are necessary for speech perception. However, perception research is very limited in what they know; and the things they claim to know are very controversial [Sch86]. So, the useful information, at least for a speech processing system, is in the feature extraction of the hearing system. The following hearing system summary gives an overview of the processing performed in the human hearing system.

The auditory system is typically divided into four main sections: outer ear, middle ear, inner ear, and central auditory nervous system [Yos85]. The outer ear consists of the (flange) ear, the external auditory canal (the ear canal), and the tympanic membrane (the ear drum). The main purpose of the outer ear is to channel sound into the middle ear. The middle ear consists of the concha, the eustachian tube, and the ossicular chain (three small bones which connect the ear drum to the inner ear). The purpose of the middle ear is to transfer changes in air pressure to vibrations against the inner ear. The auditory portion of the inner ear is the fluid filled cochlea. The cochlea transforms vibrational information into neurological information which is transmitted to the central auditory nervous system. The central auditory nervous system transfers the neurological information containing relevant speech features to the brain for further processing.

The outer ear is the easiest part of the auditory system to analyze because it is open for investigation. The outer ear's main function is to filter incoming pressure fluctuations. As with the speech production system, acoustical theory can analyze the frequency response of the outer ear. The flange's shape enhances signals within the range of frequencies 1500 Hz - 7500 Hz with its maximum gain of about 2dB at 4000 Hz [Gul89]. The flange also localizes incoming sounds. The external auditory canal's shape can be modeled as an open ended acoustical tube whose resonant wavelength equals four times the length of the tube. For typical canal lengths of 10.4cm, the resonant frequency equals 333 Hz [Mil81]. However, because the canal is somewhat soft, the resonance is distributed over three octaves. The maximum gain of the canal is about 10dB at 3333 Hz. The overall transfer function of the outer ear amplifies sound in the range of 1500 Hz to 7500 Hz, with the maximum gain at around 3000 Hz. The ear drum takes the filtered signal (in the form of pressure fluctuations) and transfers the air pressure energy to the mechanical energy system consisting of the middle ear.

The middle ear transfers air pressure information from the outer ear to mechanical vibrations in the fluid filled inner ear by way of the oval window [Yos85]. The ossicular chain connects the eardrum to the cochlea. The ossicular chain acts like an acoustical transformer which matches the impedance of the air medium of the outer ear to the fluid medium of the inner ear. The transfer function of the middle ear is in the form of a low pass filter which attenuates frequencies at about 15dB per octave for frequencies above 1000 Hz [Gul89].

The inner ear transforms the mechanical vibrations initiated by the middle ear into electrochemical excitations that are transmitted to the brain via the central auditory nervous system. The inner ear contains the cochlea, which accepts mechanical vibrations at its input (the oval window) and converts the information into electrical excitation on the auditory neural output. The cochlea is a liquid filled tube whose shape resembles a snail coiled in a spiral consisting of 2.5 turns [Yos85]. Running the full length of the cochlea,

the basilar membrane vibrates slowly (due to the surrounding incompressible fluid) in response to inputs from the middle ear. Any excitation at the oval window sets a traveling wave down the basilar membrane. However, depending on the frequency content of the excitation, the wave may not travel the entire distance of the basilar membrane; and in fact, will obtain its maximum displacement at a distance on the membrane which is directly related to the signal's frequency. Frequency selectivity of the basilar membrane occurs for two reasons. First, the basilar membrane's width changes monotonically over the length of the tube. At the oval window, the basilar membrane is at its greatest width, and at the end of the tube is its smallest width. Second, the tension of the basilar membrane is maximum at the oval window and decreases monotonically to the end of the membrane. These changing parameters of the basilar membrane cause a conversion of the input vibrations to frequency information. Low frequency excitations obtain their maximum displacement near the end of the basilar membrane while high frequency signals obtain their maximum displacement close to the oval window. Therefore, one can think of the basilar membrane as a spectrum analyzer which outputs the frequency content of the vibrational excitation as deflections at certain points along the membrane. Each point along the basilar membrane can be modelled as a band pass filter whose center frequency is a function of the point location. Also, the monotonic decrease in the basilar membranes tension is not a linear function, but an exponential function [Gul89]. Therefore, frequency is quantized logarithmically.

Along the entire length of the basilar membrane is the organ of Corti [Yos85]. The organ of Corti contains several rows of hair cells which move laterally with respect to the vibrational movement of the basilar membrane. Each hair cell contains several small hairs called stereocillia which, when deflected by the lateral movement of the hair cell, expel proteins that stimulate the neuron output. Therefore, the stereocillia transforms the vibrational information received from the basilar membrane into chemical information. There are thousands of hair cells distributed along the length of the organ of Corti and

about 30,000 neural outputs [Gei88] form the collection of hair cells. Therefore, the frequency selectivity of the basilar membrane is contained in the thousands of hair cells while information, relating the frequency decomposition to the speech signal, transmits to the brain using 30,000 neurons.

The central auditory nervous system communicates the frequency selective information generated by the inner ear to the brain. Each neuron is attached to the set of hair cells in the inner ear. A neuron represents the bandpass response of the particular section of the basilar membrane from which the neuron originated. The interesting thing about this section of the auditory system is the method in which this bandpass information is transmitted to the brain. Each neuron communicates by changing its resting voltage potential [Yos85]. Thus, by measuring the voltage of a neuron with respect to time, the communication process between the inner ear and brain can be monitored. By analyzing the information passed to the brain on the neural transmitters, one may find the features which are extracted by the auditory process from incoming sounds. Unfortunately, neurons are limited in how they can transmit information, so decoding the feature information is difficult. A neuron communicates by firing a impulse shaped voltage along its path. With respect to time, a neuron output looks like an impulse train of voltage potentials. The discharge rate (the number of times a neuron discharges or fires in a given time period) is quite limited in the auditory system to a maximum of 1,000 times per second [Gul89]. For no sound, neurons fire at an average rate between 10 and 50 times a second [Yos85]. When a sound is present, neurons respond depending on the frequency content and intensity of the sound. Therefore, the frequency selective information found by the basilar membrane is communicated to the brain in a very nonlinear system formed by impulse trains. Neurons communicate information by coding information into the impulse voltages by changing the firing rate and changing the interval time between each impulse. Exactly what the impulses represent is still not completely known [Gul89]. However, a few things have been discovered about neural

transmission. First, as the intensity of a particular frequency (especially high frequencies) increases, the firing rate of the appropriate neurons also increases [Yos85]. For low frequency tones, it has been found that the firing rate of neurons seem to track or phase-lock onto the tone. For instance, a 1000 Hz tone will produce a firing rate of approximately once per .001 seconds. Also, neurons only fire on positive portions of a signal in most cases [Yos85]. Therefore, another very nonlinear element (a clipping function) is embedded in the hearing system. In addition to these properties, the characteristics of some neurons respond more to steady state sounds while other neurons respond to changes in the frequency content of signal [Pol88]. This shows that each neuron provides a significant amount of speech information. This information is currently being studied using neurograms which are displays of neuron firings versus time [Pis85].

This discussion on the auditory system shows the hearing system's complexity and how by analyzing the hearing system a better speech processing system may be developed some day. Certainly the frequency content of speech places a major role in recognition since the inner ear decomposes incoming sounds into a series of bandpass filters spanning the frequency range of hearing. Also of interest is how the frequency scale is spaced logarithmically, not linearly. Exactly what other information the brain uses for recognition is not known, but any speech processing algorithm should, at the minimum, analyze the frequency content of the speech. Also, because the brain receives the frequency information as a function of time and because some neurons respond stronger to changes in frequency, a speech processing system should perform a frequency analysis versus time. The traditional techniques which perform a time versus frequency analysis (typically LPC analysis or spectrogram analysis) suffer however from a necessary assumption of short-time time-invariance. This is the rational for using time-frequency representations for analyzing speech signals.

VITA²

Matthew Richard Perry

Candidate for the Degree of

Doctor of Philosophy

Thesis: A CORRELATION APPROACH TO TIME-FREQUENCY
REPRESENTATIONS

Major Field: Electrical Engineering

Biographical:

Personal Data: Born in Dearborn, Michigan, July 24, 1962, the son of Richard and Carol Ann Perry. Married to Cindy Walker on May 24, 1986.

Education: Graduated from Memorial High School, Tulsa, Oklahoma, in May 1980; received the Bachelor of Science degree in Electrical Engineering from Oklahoma State University in December, 1984; received the Master of Science degree from the Oklahoma State University in May 1987; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in July, 1991.

Professional Experience: Associate Electrical Systems Engineer, E-Systems, Greenville, Texas, January, 1985 to January, 1986. Electrical Engineer, Frontier Engineering, Stillwater, Oklahoma, January, 1986 to June, 1986. Research Assistant, Department of Electrical Engineering, Oklahoma State University, July, 1986 to May, 1987. Senior Engineering Systems Analyst, Texas Instruments, Lewisville, Texas, June, 1987 to June, 1989. Research Assistant, Department of Electrical Engineering, Oklahoma State University, July, 1989 to July, 1991.